E-20-J22
#2

Multiple Pollutants and Cardiorespiratory Outcomes – Air Quality Data Analysis

Annual Progress Report to Emory University, under NIEHS contract
Period of performance: August 1, 2003 – July 31, 2004

Submitted by Amit Marmur, Katie Wade, Ted Russell, and Jim Mulholland
Georgia Institute of Technology
February 7, 2005

Summary

Air quality data and analysis are being provided to support an Emory School of Public Health research team in an epidemiologic investigation of relationships between multiple pollutants and risk of cardiac and respiratory emergency department visits in Atlanta. Emory is conducting a suite of investigations, under EPRI, EPA and NIEHS sponsorship, titled the Study of Particulates and Health in Atlanta (SOPHIA). Collaboration on the study described in this report began in September, 2002, and continues for four years. The study expands an existing study to cover the period January 1, 1993 through August 31, 2002. Several air quality databases are being used in this study. First, meteorological data are collected from the National Climatic Data Center (NCDC) network. Daily values of mean barometric pressure, maximum and minimum temperature, maximum and minimum relative humidity, total precipitation, mean dew point, average wind speed, average wind vector, and mean visibility are obtained from the Hartsfield Airport station. Second, EPA's Air Quality System (AQS) database is being used to provide ambient air monitoring data for criteria pollutants. Particle mass data are also used from the Georgia Department of Natural Resources' Metro Atlanta Index (MAI) database. Third, the database from Atlanta's Jefferson St. superstation for the time period July 1, 1998 through December 31, 2002, developed by Atmospheric Research & Analysis, Inc., under Electric Power Research Institute sponsorship as a part of the SouthEastern Aerosol Research and Characterization (SEARCH) Study, is being used. Fourth, data from the Assessment of the Spatial Aerosol Composition in Atlanta (ASACA), which began March 1, 1999, and is directed by co-PI Ted Russell, are being used in this study. The SEARCH and ASACA databases contain detailed information about the size and composition of airborne particles.

This report summarizes three areas of research by Georgia Tech researchers undertaken during the second year to support the epidemiologic study Multiple Pollutants and Cardiorespiratory Outcomes, sponsored by NIEHS. First, the comprehensive air quality database has been extended through December 31, 2002. Second, temporal and spatial analyses have been performed to maximize the precision and accuracy of the data and completeness of the dataset, to assess the representativeness of monitoring sites, and to estimate error associated with the measures. Third, source apportionment analysis of daily $PM_{2.5}$ measures has been performed using the chemical mass balance approach. Papers drafted on the second and third topics are appended.

Development of Air Quality Database

The air quality database containing measures of criteria pollutants has been extended through December 31, 2002. The monitoring sites included are listed in Table 1 below, and locations are shown in Figure 1 on the next page. Hourly values of the following pollutant gases were obtained: sulfur dioxide (SO_2), carbon monoxide (CO), nitrogen dioxide (NO_2), nitrogen oxides (NO_x), and ozone (O_3). Daily and hourly measures of particulate matter mass, both PM_{10} and $PM_{2.5}$, were obtained.

Table 1. Air quality monitoring stations in metro-Atlanta used in this study.

station	*measure(s)*	*longitude (W)*	*latitude (N)*
Hartsfield Int. Airport (HA) [a]	meterological data	-84.433	+33.633
Bolton Rd (BR) [a]	precipitation	-84.500	+33.833
Fulton County (FC) [a]	visibility	-84.517	+33.783
Roswell Rd (RR) [b]	CO	-84.3803	+33.8764
Dekalb Tech (DT) [b]	CO	-84.2358	+33.7892
Confederate Ave. (CA) [b]	SO_2, O_3	-84.3578	+33.7206
Stilesboro (St) [b]	SO_2	-84.9153	+34.1033
Georgia Tech (GT) [b]	NO_2, NO_x, SO_2, PM_{10}	-84.4008	+33.7758
S. Dekalb College (SD) [b,d]	NO_2, NO_x, O_3, $PM_{2.5}$	-84.2903	+33.6875
Conyers Monastery (Co) [b]	NO_2, NO_x, O_3	-84.0667	+33.5856
Tucker (Tu) [b,d]	NO_2, $PM_{2.5}$	-84.2142	+33.8472
Yorkville (Yo) [b,e]	CO, SO_2, O_3, NO_2, NO_x, PM_{10}, $PM_{2.5}$	-85.0453	+33.9283
Fire Station # 8 (FS8) [b]	PM_{10}, $PM_{2.5}$	-84.4358	+33.8017
Fulton County Health Dept. (FCHD) [b,c]	PM_{10}	-84.3828	+33.7517
Doraville Health Center (DHC) [b]	PM_{10}	-84.2789	+33.9031
Griffin (Gr) [b]	PM_{10}	-84.2850	+33.2647
Douglasville (Do) [b]	$PM_{2.5}$	-84.7789	+33.7433
E. Rivers School (ERS) [b]	$PM_{2.5}$	-84.3819	+33.8137
East Point Health Center (EPHC) [b]	$PM_{2.5}$	-84.4375	+33.6164
Forest Park (FP) [b]	$PM_{2.5}$	-84.3911	+33.6097
Kennesaw (Ke) [b]	$PM_{2.5}$	-84.6075	+34.0144
Fort McPherson (FM) [d]	$PM_{2.5}$	-84.4375	+33.7083
Jefferson St. (JS) [d,e]	CO, SO_2, O_3, NO_2, NO_x, PM_{10}, $PM_{2.5}$	-84.4167	+33.7769

[a] NCDC; [b] AQS; [c] MAI; [d] ASACA; [e] SEARCH

Hourly data files have been archived. When a small number of hourly data are missing, values are interpolated. Then, daily values with various averaging times are calculated. Files containing daily values as well as temporal and spatial patterns are listed below. These have been put on a web site for use by the research teams at Georgia Tech and Emory.

meteorological data:	MET.xls
SO_2 data:	SO2.xls
CO data:	CO.xls
NO_2/NO_x data:	NO2-NOx.xls
O_3 data:	O3.xls
PM_{10} data:	PM10.xls
$PM_{2.5}$ data:	PM25.xls

As a part of a complimentary EPA study, a program has been written to identify outliers. Composite values have also been computed.

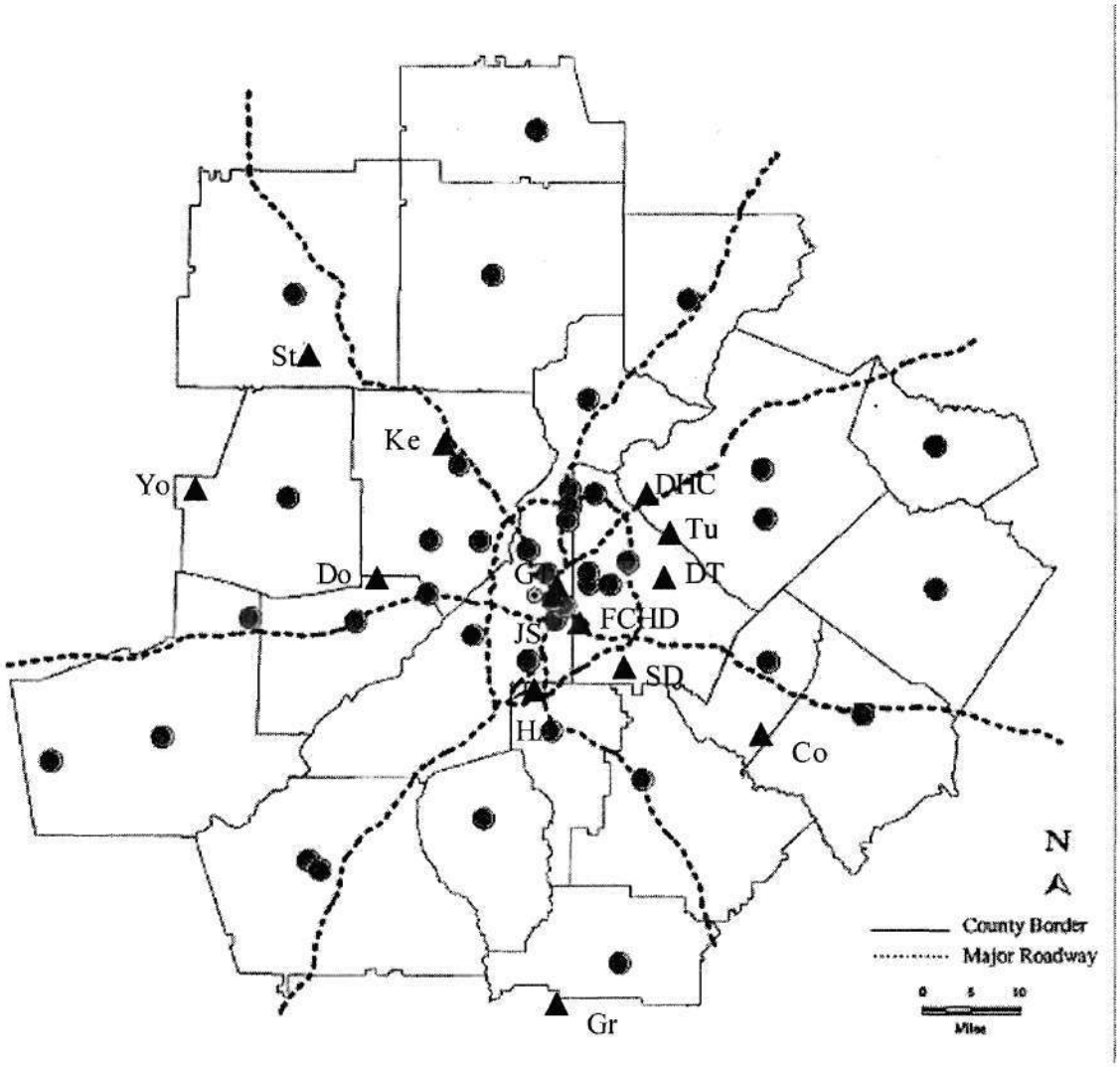

Figure 1. Location of selected air quality monitoring stations (triangles). Also shown are the locations of hospitals participating in the study (circles).

Averages and ranges for air quality data measures are given in Table 2. Temporal and spatial trends have been described as a part of the EPA study. For the NIEHS study described here, we are performing analyses that address the representativeness of the Jefferson St. (JS) central monitoring station and error associated with air pollutant measures. PM source apportionment using receptor-based chemical mass balance modeling is also being performed.

Appendix A (page 5): Measurement Error and Spatial Variability of Ambient Air Pollution in Atlanta, Georgia

Appendix B (page 25): Optimization Based Source-Apportionment of PM2.5 Incorporating Gas-to-Particle Ratios

Table 2. Daily air quality measures, January 1, 1993 through December 31, 2002 (3652 days).

	station	*frequency*	*average (range)*
MET: mean pressure	HA	daily	981 millibar (959-1001)
maximum temperature	HA	daily	72.4 °F (18 – 102)
minimum temperature	HA	daily	53.3 °F (6 – 80)
maximum relative humidity	HA	daily	88.3 % (40 – 100)
minimum relative humidity	HA	daily	48.0 % (7 – 100)
mean dew point	HA	daily	50.7 °F (-2.4 – 74.7)
total precipitation	HA,BR	daily	0.13 inches (0 – 6.7)
mean wind speed	HA	daily	8.1 mph (0.3 – 28.1)
mean wind vector	HA	daily	1.4 mph (0.1 – 23.9); 313°
mean visibility	HA,FC	daily	9.3 miles (0.3 – 21.3)
SO_2 (1-hr max; 24-hr avg)	GT	hourly	15.3 ppb (1 – 149)
	CA	hourly	12.5 ppb (1 – 136)
	JS	hourly, since 8/1/98	18.6 ppb (0.05 – 101)
	St	hourly	12.4 ppb (1 – 273)
	Yo	hourly	11.6 ppb (0.2 – 158)
CO (1-hr max; 8-hr max; 24-hr avg)	DT	hourly	1.68 ppm (0.1 – 7.7)
	RR	hourly, since 8/5/94	2.00 ppm (0.2 – 9.5)
	JS	hourly, since 8/1/98	1.37 ppm (0.20 – 7.69)
	Yo	hourly	0.30 ppm (0.08-1.05)
NO_2 (1-hr max; 24-hr avg)	GT	houly	44.6 ppb (7 – 181)
	JS	hourly, since 8/1/98	52.9 ppb (0.1 – 200)
	SD	hourly	33.4 ppb (1 – 139)
	Tu	hourly, since 4/6/95	32.7 ppb (1 – 100)
	Co	hourly, since 4/1/94	14.7 ppb (1 – 80)
	Yo	hourly, since 1/25/96	11.7 ppb (1 – 70)
NO_x (1-hr max; 24-hr avg)	GT	hourly, since 4/1/98	133 ppb (5 – 996)
	JS	hourly, since 8/1/98	153 ppb (6 – 671)
	SD	hourly, since 4/1/94	158 ppb (1 – 930)
	Tu	hourly, since 4/6/95	64.7 ppb (1 – 387)
	Co	hourly, since 4/1/94	27.2 ppb (1 – 187)
	Yo	hourly, since 1/25/96	13.8 ppb (1 – 148)
O_3 (1-hr max; 8-hr max; 24-hr avg)	CA	hourly (93-95: 3/1–11/30)	54.0 ppb (3 – 147)
	SD	hourly (96: 3/1–10/31)	49.7 ppb (2 – 138)
	Co	hourly (97-02: 4/1–10/31)	54.5 ppb (5 – 133)
	JS	hourly, since 8/1/98	44.3 ppb (1 – 131)
	Yo	hourly	51.5 ppb (3 – 133)
PM_{10} mass (24-hr avg)	GT	daily, since 1/9/96	26.0 $\mu g/m^3$ (4 – 98)
	JS	daily, since 7/15/98	27.8 $\mu g/m^3$ (4 – 100)
	FCHD	5 dy/wk, 1/1/93–1/3/96	29.8 $\mu g/m^3$ (0 – 151)
	FS8	6th day	32.0 $\mu g/m^3$ (0 – 80)
	DHC	6th day	24.8 $\mu g/m^3$ (0 – 74)
	Yo	daily, since 7/25/98	21.2 $\mu g/m^3$ (5 – 62)
	Gr	6th day	23.4 $\mu g/m^3$ (0 – 82)
$PM_{2.5}$ mass (24-hr avg)	JS	daily, since 7/10/98	18.0 $\mu g/m^3$ (3 – 66)
	FM	daily, since 3/1/99	19.2 $\mu g/m^3$ (3 – 75)
	Tu	daily, since 3/1/99	19.8 $\mu g/m^3$ (3 – 52)
	SD	daily, since 1/1/99	18.1 $\mu g/m^3$ (5 – 47)
	DHC	daily, since 1/1/99	18.4 $\mu g/m^3$ (0.7 – 89)
	ERS	daily, since 1/1/99	18.0 $\mu g/m^3$ (2 – 140)
	FS8	3rd day, since 1/1/99	20.1 $\mu g/m^3$ (0.8 – 79)
	EPHC	3rd day, since 1/1/99	18.3 $\mu g/m^3$ (0.4 – 77)
	FP	3rd day, since 1/1/99	18.2 $\mu g/m^3$ (0.7 – 74)
	Ke	3rd day, since 1/1/99	18.0 $\mu g/m^3$ (2 – 94)
	Yo	3rd day, since 5/6/98	14.3 $\mu g/m^3$ (2 – 65)

Appendix A:

Measurement Error and Spatial Variability of Ambient Air Pollution in Atlanta, Georgia

Draft manuscript to be submitted to the *Journal of the Air & Waste Management Association*

Measurement Error and Spatial Variability of Ambient Air Pollution in Atlanta, Georgia

Katherine S. Wade, James A. Mulholland, Amit Marmur, and Armistead G. Russell
School of Civil & Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia
Ben Hartsell and Eric Edgerton
Atmospheric Research & Analysis, Inc., Durham, North Carolina
Mitch Klein, Lance Waller, Jennifer L. Peel and Paige E. Tolbert
Rollins School of Public Health, Emory University, Atlanta, Georgia

ABSTRACT

Data from the U.S. Environmental Protection Agency's Air Quality System (formerly known as Aerometric Information Retrieval System), the SouthEastern Aerosol Research and Characterization database, and the Assessment of Spatial Aerosol Composition in Atlanta database for 1999 through 2002 have been used to characterize measurement error and the spatial variability of various ambient air pollutants in Atlanta, Georgia. These data are being used in time-series epidemiologic studies in which associations of acute respiratory and cardiovascular health outcomes and daily ambient air pollutant levels are assessed. Normalized semivariograms are used to quantify the effects of exposure assessment error for gaseous pollutants (SO_2, CO, NO_x, and O_3) and fine particulate matter ($PM_{2.5}$ mass, sulfate, nitrate, ammonium, elemental carbon, and organic carbon). For the gaseous pollutants and $PM_{2.5}$ mass, instrument error represents 5 to 10% of the temporal variation in the pollutant. For $PM_{2.5}$ constituents, measurement error ranged from 10 to 40% of the temporal variation. Spatial variability is greatest for primary pollutants (SO_2, CO, NO_x, and elemental carbon). Wind rose plots are used to demonstrate the effects of point sources and local (road) sources on monitoring station data. The results presented here will help quantify the impact of measurement error and spatial variability on the assessment of health effects of ambient air pollution in Atlanta and are relevant for interpreting the use of the data from other fixed monitors in health related studies.

INTRODUCTION

In epidemiologic time-series studies in which the short-term health effects of ambient air pollution are assessed, measurement error and the spatial variability of air pollution can impact the assessment. A number of studies have addressed the limitations of using central monitoring station data as exposure measurements[1-4], though, in population-based epidemiologic studies, while differences between actual exposures and central station values can be large, the daily mean of personal exposures is likely to be better correlated with a central station value than an individual exposure level [5]. Introduction of such uncertainty in the exposure measurement is important because it tends to reduce the ability of epidemiologic studies to assess the health effects of air pollution, decreasing the strength of association estimate (bias to null) and increasing the width of its associated confidence interval. The attenuation varies across pollutants because measurement error and spatial variation of air pollutants differ. In this paper, we address the issue of exposure misclassification errors in ecologic time-series studies of air pollution and health in Atlanta, Georgia.

Results of the Aerosol Research Inhalation Epidemiology Study (ARIES) of emergency department visits for cardiovascular and respiratory diseases in relation to ambient air pollution in Atlanta, Georgia, from 1993 to 2000 have recently been published[6,7]. For cardiovascular

disease, positive associations were observed with nitrogen dioxide (NO_2), carbon monoxide (CO), fine particulate matter ($PM_{2.5}$), and $PM_{2.5}$ components organic carbon and elemental carbon. Risk ratios ranged from 1.04 to 1.02 per standard deviation increase for these pollutant measures. For respiratory disease, positive associations were observed with ozone (O_3), NO_2, CO, and PM_{10}. Central monitoring station data were used in this assessment, including data obtained since August, 1998, from the Southeastern Aerosol Research and Characterization Study (SEARCH) site near downtown Atlanta at Jefferson Street[8]. Characteristics of ambient air pollution data that affect interpretation of the epidemiologic results include instrument error, local source effects, and spatial heterogeneity due to factors such as meteorological phenomena, topological features, and pollutant volatility and reactivity.

Instrument error can be quantified from co-located instrument data. For continuous and semicontinuous measurements, instrument error typically results from calibration drift, flow rate changes, and changes in atmospheric conditions such as relative humidity. These errors are minimized through stringent qua lity control protocols. For filter-based measurements, such as PM mass, ions, organic carbon and elemental carbon, instrument error also results from sample handling and laboratory analysis.

A number of studies have examined the spatial variability of air pollutants in urban areas. When data from a sufficient number of monitoring stations are available, spatial interpolation techniques can be used to provide a spatially continuous representation. In previous work, we used a universal kriging procedure for estimating daily ozone concentrations for each zip code in the Atlanta metropolitan statistical area (MSA) [9]. However, the area of representativeness for a given monitor for a given pollutant can lead to uncertainty in a study[10,11]. Several studies have used Pearson correlation coefficients to characterize spatial variability in ambient air pollutant concentrations[12-16]. A plot of the correlation coefficient versus distance between two sites provides a measure of spatial autocorrelation. Alternatively, the semivariogram, which uses the covariance instead of the correlation, is a well-established tool for conveying information about the spatial variability of environmental pollutants [17,18]. Diem[19] has argued against the use of semivariogram analysis for air pollution as there are often not enough monitoring stations available to provide a suitable number of points to plot. In particular, the Atlanta MSA would need 100 monitoring stations to create a stable semivariogram of ozone. Obviously, this is an unreasonable expectation. Although Diem suggests that spatial modeling can be used to surmount this obstacle, that is not the point of the current study. The authors instead wish to provide a metric that can be used within existing monitoring networks to assess the spatial accuracy of these networks when applied to continuous study areas, for example in an epidemiological study over a city. In such instances, the exact point of exposure is often not known and instead a central or averaged value is used. The semivariograms developed in this study can be used to qualify the accuracy of this central value.

Although ambient air quality monitoring stations are sited to minimize local source effects, impacts of specific point and roadway sources are observed, particularly for primary pollutants. For example, Duncan et al.[20] showed the contribution of power plant plume fumigation events on nitrogen oxides ($NO_x = NO + NO_2$) and sulfur dioxide (SO_2) concentrations in Atlanta. Kirby et al. showed the contribution of roadways to nitrogen oxides.[21] Secondary air pollutants, such as O_3 and a significant fraction of $PM_{2.5}$ mass, exhibit high spatial autocorrelation. From analyses of the spatial variability of $PM_{2.5}$ in several urban areas in the southeast, Pinto et al.[22] found high correlations between site pairs and spatial uniformity in concentration fields. In a study of spatial aerosol composition in Atlanta in 1999,

Butler et al.[23] concluded that $PM_{2.5}$ mass and chemical composition were relatively spatially homogeneous.

In this paper, we address the following two questions. First, for each ambient air pollutant used in time-series studies of the short-term health effects of air pollution in Atlanta, what is the error in the exposure variable due to measurement error and spatial variability relative to the temporal variation in this exposure variable that provides the power with which to identify an association? Second, to what extent do local and specific point sources impact measurements at each of the ambient air quality monitoring stations in Atlanta?

METHODS

Ambient Air Quality Monitoring Stations

Ambient air quality data from the U.S. Environmental Protection Agency's Air Quality System (AQS, formerly known as Aerometric Information Retrieval System), the SEARCH database, and the Assessment of Spatial Aerosol Composition in Atlanta (ASACA) are used to characterize measurement error and spatial variability of air pollution in the Atlanta MSA for the four-year period 1999-2002. Monitoring site locations are shown in Figure 1. The Jefferson Street site is of particular focus here because it was the location of the Atlanta Supersite, and Atlanta-based epidemiological studies rely heavily on data from that location. Pollutants used in this study, for which daily measures were available at multiple sites in the Atlanta MSA, are four pollutant gases (SO_2, CO, NO_x, and O_3), $PM_{2.5}$ total mass, and five $PM_{2.5}$ components (sulfate ion (SO_4^{2-}), nitrate ion (NO_3^-), ammonium ion (NH_4^+), elemental carbon (EC), and organic carbon (OC)). The completeness of the data set was very high (>90% for most pollutants). Average values and correlation coefficients with data from the Jefferson Street (JS) monitoring station near downtown Atlanta are listed in Tables 1 and 2.

Sulfur dioxide data from five monitoring sites in the Atlanta MSA were available. Located near downtown is the SEARCH monitor at Jefferson Street and AQS monitors at Georgia Tech (GT) and Confederate Avenue (CA). Located approximately 60 km to the northwest are AQS monitors at Stilesboro (St) and Yorkville (Yo). For the health effect studies, the one-hour maximum of SO_2 was used. SO_2 levels are low in Atlanta, on average between 3 and 6 ppb. Levels at Jefferson Street were higher than the other sites, possibly due to differences in the analytical method of drying the gas stream prior to analysis.

Carbon monoxide data from four monitoring stations were available. In addition to the Jefferson Street site, CO monitors are located at Roswell Road (RR), Dekalb Tech (DT), and Yorkville (Yo). Highest CO levels are observed at Roswell Road, which is located near high traffic density roads.

Nitrogen oxides are measured at six sites: Jefferson Street (JS), Georgia Tech (GT), South Dekalb (SD), Tucker (Tu), Conyers (Co) and Yorkville (Yo). Within Atlanta's perimeter highway, average NO_x levels exceed 40 ppb. NO_x levels are significantly lower at the rural Conyers and Yorkville sites.

Ozone data from five stations were used: Jefferson Street (JS), Confederate Avenue (CA), South Dekalb (SD), Conyers (Co), and Yorkville (Yo). Three daily measures of O_3 are used: one-hour maximum, eight-hour maximum, and 24-hour average. During the period 1999 through 2002, AQS sites only reported data from April through October. Ozone levels are spatially very uniform and highly correlated.

Fine particulate matter mass data from eight AQS stations, two SEARCH sites, and three ASACA sites were used. AQS stations at E. River School (ERS), South Dekalb (SD), and Doraville Health Center (DHC) monitor daily $PM_{2.5}$ mass by the filter-based FRM (Federal Reference Method). AQS stations at Fire Station #8 (FS8), East Point Health Center (EPHC), Forest Park (FP), Kennesaw (Ke), and Yorkville (Yo) report $PM_{2.5}$ mass by FRM every three days. At the ASACA sites at South Dekalb (SD), Tucker (Tu), and Fort McPherson (FM), $PM_{2.5}$ mass is measured continuously by TEOM (Tapered Element Oscillating Microbalance). At the SEARCH sites at Jefferson Street and Yorkville, $PM_{2.5}$ mass is measured both by FRM and TEOM.

$PM_{2.5}$ major ion (SO_4^{-2}, NO_3^-, and NH_4^+) and carbon fraction (EC and OC) data from March 1999 through August 2000 at five locations in the Atlanta MSA were used (Table 2). Filter-based PCMs (Particle Composition Monitors) were used that included three channels to collect 24-hour integrated samples for analysis of major ions, trace metals, organic and elemental carbon in $PM_{2.5}$ size range. Ion chromatography was used to quantify water-soluble ionic species. Elemental and organic carbon collected on quartz filters were measured by Thermal Optical Transmittance (TOT) in the ASACA network (FM, SD, and Tu), and by Thermal Optical Reflectance (TOR) in the SEARCH network (JS and Yo). Comparison of these two techniques indicates that while the total carbon measurements are in good agreement, lower EC and higher OC values are obtained using TOT compared to TOR[24].

Spatial Statistical Analysis

Using methods from the field of geostatistics, we develop an approach for describing the spatial autocorrelation of air pollutant data. Consider spatial exposure data $Z(s_1)$, $Z(s_2)$, …, $Z(s_N)$ that represent daily observations of air pollutant Z at monitoring site locations s_1, s_2, …, s_N. The daily data at a given site are assumed to be randomly distributed with mean μ and standard deviation σ. For a pair of monitoring sites at locations s_i and s_j, the correlation coefficient R is defined as follows.

$$R(h) = \frac{1}{n}\sum \frac{(Z(s_i) - \mu_i)(Z(s_j) - \mu_j)}{\sigma_i \sigma_j} \qquad (1)$$

Here, n is the number of daily observations and h is the distance between s_i and s_j. Spatial autocorrelation is an attribute of spatial data based on the fact that observations closer together tend to be more alike than observations farther apart. If the spatial process is isotropic, then R is a function of distance alone (independent of direction). The correlogram plot is a graph of R versus the distance h. For small distances, R is close to one; as distance increases, R decreases, approaching zero as the spatial correlation approaches zero.

All data is initially log-normalized. In our population-based epidemiologic studies of the short-term health effects of air pollution in Atlanta, the daily variation in air pollution drives the health risk assessment. We therefore also normalize the data at each monitoring site in the Atlanta MSA as follows.

$$Z'(s) = \frac{Z(s) - \mu}{\sigma} \qquad (2)$$

The distributions of normalized data Z' are the same at each site, with a mean of zero and a standard deviation of one. This normalization allows EC and OC from the TOT and TOR to be used together as it removes the bias between the two approaches, assuming that spatial inhomogeneity in the sources does not lead to an added apparent spatial variability in the data.

Given the strong correlation found between the two techniques[25], this is likely the case. The correlation coefficient is the same for the actual observations Z and the normalized data Z'.

Another measure of spatial autocorrelation is the semivariogram[26]. The semivariogram (γ) is one-half the variogram:

$$\gamma(s_i - s_j) = \frac{1}{2} Var(Z'(s_i) - Z'(s_j)) \quad (3)$$

It can be shown that the semivariogram (eq 3) is related to the correlogram as follows.

$$\gamma(h) = 1 - R(h) \quad (4)$$

Thus, the graph of this semivariogram versus distance is an inverted form of the correlogram, starting near zero at short distances and increasing with distance, approaching one if the spatial correlation approaches zero at large distances. In the geostatistical field, the shape of the semivariogram is described by the nugget, which refers to a non-zero semivariogram near the origin, the range, which refers to the separation distance at which the semivariogram levels off, and the sill, which refers to the value of the semivariogram at which it levels off.

In this paper, we propose a measure of spatial autocorrelation that can most easily be interpreted in terms of potential effect of measurement error and spatial variability on the epidemiologic study findings. That is, we compute the ratio of the standard deviation of the spatial variation from the average of two sites (spatial sd) to the standard deviation of the temporal variation of the average of two sites (temperal sd). Here, the spatial variation in an air pollutant concentration represents exposure error, and the temporal variation of the daily measures represents the power with which health risk can be assessed. The daily spatial average of the normalized variable Z' at two monitoring sites s_i and s_j is computed as follows.

$$\overline{Z'} = \frac{Z'(s_i) + Z'(s_j)}{2} \quad (5)$$

Analogous to the semivariogram, it can be shown that the spatial sd and the temporal sd are related to the correlation coefficient as follows.

$$\text{spatial sd} = \sqrt{\frac{1 - R(h)}{2}} \quad (6)$$

$$\text{temporal sd} = \sqrt{\frac{1 + R(h)}{2}} \quad (7)$$

Thus,

$$\text{spatial sd / temporal sd} = \sqrt{\frac{1 - R(h)}{1 + R(h)}} \quad (8)$$

A plot of the ratio given by eq 8 versus distance between monitoring stations has the same general shape as the semivariogram plot. This normalized semivariogram can be interpreted as the fraction of temporal variation in a pollutant that is lost due to exposure assessment error. The nugget can be interpreted as the effect of instrument error, as measured by data obtained from co-located instruments for most pollutants.

RESULTS AND DISCUSSION

Measurement Error and Spatial Variation Effects

Normalized semivariograms for one-hour maximum CO, one-hour maximum NO_x, one-hour maximum SO_2, eight-hour maximum O_3, and 24-hour average $PM_{2.5}$ mass are shown in

Figure 2. As expected, the primary pollutants (CO, NO_x, and SO_2) exhibited much greater spatial variability than the predominantly secondary pollutants (O_3 and $PM_{2.5}$). That is, the sill of the primary pollutant semivariograms appears to have a value of one, indicating that, at large distance (100 km), the uncertainty in the exposure variable due to spatial variability of the primary pollutant is as large as the temporal variation in that pollutant. The sills for O_3 and $PM_{2.5}$, on the other hand, appear to be approximately 0.3 and 0.4, respectively. Since concentrations of these pollutants are not driven by specific sources but by regional meteorology and chemistry, spatial variation for the secondary pollutants accounts for a smaller fraction of temporal variation than for the primary pollutants.

Measurement error, as assessed by co-located instrument data obtained from the ambient air monitoring program of the Georgia Department of Natural Resources and SEARCH network, ranged from 5% of the temporal variation for O_3 to 10% for $PM_{2.5}$ mass. The higher value of the nugget for $PM_{2.5}$ mass is likely due to laboratory error (e.g., humidity control) associated with the filter-based $PM_{2.5}$ mass measurements[27]. The range in the semivariogram was smallest for SO_2. As will be shown later, this is likely due to the local nature of power plant plume fumigation events that impact SO_2 measurements at the ambient monitoring stations.

Normalized semivariograms for 24-hr integrated measurements of $PM_{2.5}$ major ions (SO_4^{-2}, NO_3^-, and NH_4^+) and carbon fractions (EC and OC) are shown in Figure 3. Consistent with the findings cited above, the semivariogram sills for the predominantly secondary components of $PM_{2.5}$ (SO_4^{-2}, NO_3^-, NH_4^+, and OC) were lower than the sill for the primary $PM_{2.5}$ component (EC). Also not surprising is the higher nugget effect for the $PM_{2.5}$ component measurements. The analytical methods used for these measurements are more involved than those used for the gases and $PM_{2.5}$ total mass, with more opportunity for measurement error.

The semivariograms (Figures 1 and 2) indicate that spatial variability of the exposure variables significantly reduces the power of health risk assessment. Moreover, the impact of exposure assessment error varies significantly between pollutants. For example, the data suggest that attenuation of health risk assessment is likely to be much greater for associations with SO_2 than with O_3. The population in the Atlanta-based epidemiologic studies is predominantly within a radius of 30 km of downtown Atlanta. For SO_2, measurement error and spatial variability at 30 km result in an uncertainty of over 80% of the temporal variation, making it very difficult to observe a small health risk. For O_3, on the other hand, uncertainty in the exposure variable at 30 km is only about 20% of the temporal variation. At this distance, the major portion of the uncertainty appears to be due to spatial variability due to local source impacts rather than instrument error. No one station appears to be an outlier in the semivariograms, suggesting that local and regional influences play similar roles for each site.

Point and Local Source Effects

To assess the impact of point sources and local road sources on the ambient air pollutant monitors, wind rose plots were constructed for those pollutants for which hourly data are available (Figures 4 through 8). Hourly measurements of pollutants and wind direction were used for the four-year period 1999-2002. Wind data bins were 12 degrees. Over 35,000 data points were possible; data completeness was approximately 90%. Because both wind direction and pollutant concentrations exhibit diurnal and seasonal patterns, effects of these patterns were controlled for in the "corrected" plots by dividing the data into 12 months and 12 two-hour periods. After correction, high concentrations should point in the direction of sources that impact the station.

In Figure 4, wind rose plots at Jefferson Street monitoring station are shown for CO, NO_x, SO_2, O_3, $PM_{2.5}$ mass, and black carbon (BC). The latter was measured by aethalometer and should correlate approximately with EC[28]. The plots of CO, NO_x and BC are similar. Possible sources of these pollutants are roadways to the south and west, a trucking facility to the north, and a bus maintenance facility to the south. The peak in these pollutants to the northeast is consistent with the alignment of a major interstate (I-85).

The plots of O_3 and $PM_{2.5}$ mass do not show strong effects of wind direction, as expected. Secondary pollutants are less affected by emission sources than primary pollutants. We do note, however, that O_3 minima are observed in directions where NO_x peaks occur. This is likely due to ozone inhibition by radical scavenging and the titration of O_3 by NO, indicating NO_x inhibition.

The SO_2 wind rose plot at Jefferson Street has a large peak when winds come from the northwest, and smaller peaks when winds come from the southwest and north. The peak corresponding to the winds from the north may be due to the trucking facility. Analysis of the peak SO_2 concentrations in the 0-12 degree category indicates the peaks occur most frequently in the morning when activity at the trucking facility is greatest. The northwest and southwest peaks typically occur in late afternoon. As shown in Figure 5, these peaks are likely due to coal-fired power plant fumigation events.

In Figure 5, SO_2 wind rose plots at five monitoring stations are shown. In the case of the Georgia Tech station at which wind data were not available, wind measurements at Jefferson Street were used. Also shown are the locations of coal-fired power plants relative to each station. As evidenced by the alignment of SO_2 peaks at each monitoring site with the direction of the power plants, the plots clearly show the impact of plumes from power plants located up to 70 km from Atlanta on SO_2 concentrations. The power plant plumes are injected above the atmospheric mixed layer and can be downwardly mixed during the peak temperature hours of the day. Plant Bowen, located 60 km northwest of Atlanta, is the largest of the coal-fired power plants; Plant McDonough, located much closer to downtown Atlanta and also to the northwest, is the smallest of these sources.

CO and NO_x wind rose plots using data from several monitoring stations are shown in Figures 6 and 7. The observed peaks in these pollutants are consistent with the directions of major roadways. The data show that none of the primary pollutant monitoring stations in Atlanta are immune to local source impacts. In the case of NO_x at South Dekalb the large peak to the north is due to close proximity of the interstate perimeter highway (I-285) in that direction.

Finally, O_3 wind rose plots using data from four monitoring stations are shown in Figure 8. These plots demonstrate minimal local source impacts on ozone, as expected. As already mentioned, NO_x inhibition likely accounts for the shape of the O_3 wind rose plot at the Jefferson Street station near downtown Atlanta. Given the large NO_x peak associated with winds from I-285 north of South Dekalb, one might have expected more NO_x inhibition in the O_3 wind rose plot at SD. One explanation is that this more suburban site is less NO_x-inhibited than more urban Jefferson Street site due greater biogenic VOC emissions at South Dekalb.

CONCLUSIONS

Measurement error and spatial variability of ambient air pollutant measurements in Atlanta have been assessed. As expected, secondary pollutants are much more spatially homogeneous and correlated than primary pollutants. Measurement error was greatest for the analysis of the carbon fractions of $PM_{2.5}$. Local source effects on primary pollutant levels were observed at each of the Atlanta monitoring stations, suggesting that spatial average values might

be better exposure measures than any single central monitoring station data in population-based epidemiologic studies of the health effects of air pollution. The relative amounts of exposure variable error need to be considered in interpreting the epidemiologic findings.

ACKNOWLEDGEMENTS

This work was supported by grants from the Electric Power Research Institute (W03253-07), the U.S. Environmental Protection Agency (R82921301-0), and the National Institute of Environmental Health Sciences (R01ES11294). This research used air quality data from a monitoring station operated by ARIES and managed by Ron Wyzga and Alan Hansen of EPRI.

REFERENCES

1. Goldstein, I.F. and Landovitz, L. Analysis of air pollution patterns in New York City Em Dash 2. Can one aerometric station represent the area surrounding it? *Atmos. Environ.* **1977**, *11*, 53-57.
2. Lipfert, F.W. and Wyzga, R.E. Air pollution and mortality: The implications of uncertainties in regression modeling and exposure measurement. *J. Air & Waste Manage. Assoc.* **1997**, *47*, 517-523.
3. Zeger, S.L.; Thomas, D.; Dominici, F.; Samet, J.M.; Schwartz, J.; Dockery, D.; Cohen, A. Exposure measurement error in time-series studies of air pollution: concepts and consequences. *Environ. Health Perpectives* **2000**, *108*, 419-426.
4. McNair, L.A.; Harley, R. A.; Russell, A.G. Spatial inhomogeneity in pollutant concentrations, and their implications for air quality model evaluation. *Atmos. Environ.* **1996**, *30*, 4291-4301.
5. Schwartz, J.; Dockery, D.W.; Neas, L.M. Is daily morality associated specifically with fine particles? *J. Air & Waste Manage. Assoc.* **1996**, *46*, 927-939.
6. Metzger, K.B.; Tolbert, P.E.; Klein, M.; Peel, J.L.; Flanders, W.D.; Todd, K.; Mulholland, J.A.; Ryan, P.B.; Frumkin, H. Ambient air pollution and cardiovascular emergency department visits. *Epidemiology* **2004**, *15*, 46-56.
7. Peel, J.L.; Tolbert, P.E.; Klein, M.; Metzger, K.B.; Flanders, W.D.; Todd, K.; Mulholland, J.A.; Ryan, P.B.; Frumkin, H. *Epidemiology*, in press.
8. Hansen, D.A.; Edgerton, E.S.; Hartsell, B.E.; Jansen, J.J.; Kandasamy, N.; Hidy, G.M.; Blanchard, C.L. The southeastern aerosol research and characterization study: part 1 – overview. *J. Air & Waste Manage. Assoc.* **2003**, *53*, 1460-1471.
9. Mulholland, J.A.; Butler, A.J.; Wilkinson, J.; Russell, A.G.; Tolbert, P.E. Temporal and spatial distributions of ozone in Atlanta: regulatory and epidemiologic implications. *J. Air & Waste Manage. Assoc.* **1998**, *48*, 418-426.
10. Tiles, S.; Zimmerman, J. Investigation on the spatial scales of the variability in measured near-ground ozone mixing ratios. *Geophysical Research Letters.* **1998**, *25*, 3827-3830.
11. Villasenor, R.; Ortiz, E.; Watson, J.; Chow, J. Spatial and temporal variations in ambient $PM_{2.5}$ and PM_{10} in Mexico City. *J. Aerosol Sci.* **2000**, 31, S901-S902.
12. Monn, C.; Carabias, V.; Junker, M.; Waeber, R.; Karrer, M.; Wanner, H.U. Small-scale spatial variability of particulate matter < 10 μm (PM_{10})and nitrogen dioxide. *Atmos. Environ.* **1997**, *31*, 2243-2247.
13. Buzorius, G.; Hameri, K.; Pekkanen, J.; Kulmala, M. Spatial variation of aerosol number concentration in Helsinki city. *Atmos. Environ.* **1999**, *33*, 553-565.

14. Morawska, L.; Vishvakarman, D.; Mengersen, K.; Thomas, S. Spatial variation of airborne pollutant concentrations in Brisbane, Australia and its potential impact on population exposure assessment. *Atmos. Environ.* **2002**, *36*, 3545-3555.
15. Lin, T.; Young, L.; Wang, C. Spatial variations of ground level ozone concentrations in areas of different scales. *Atmos. Environ.* **2001**, *35*, 5799-5807.
16. Rao, S.T.; Zalewsky, E.; Zurbenko, I.G.; Determining temporal and spatial variations in ozone air quality. *J. Air & Waste Manage. Assoc.* **1995**, *45*, 57-61.
17. Grondona, M.O. and Cressie, N. Using spatial considerations in the analysis of experiments. *Technometrics* **1991**, *33*, 381-392.
18. Casado, L.S.; Rouhani, S.; Cardelino, C.A.; Ferrier, A.J. Geostatistical analysis and visualization of hourly ozone data. *Atmos. Environ.* **1994**, *28*, 2105-2118.
19. Diem, J.E. A critical examination of ozone mapping from a spatial-scale perspective. *Environ. Pollution* **2003**, *125*, 369-383.
20. Duncan, B.N.; Stelson, A.W.; Kiang, C.S. Estimated contribution of power plants to ambient nitrogen oxides measured in Atlanta, Georgia in August 1992. *Atmos. Environ.* **1995**, *29*, 3043-3054.
21. Kirby, C.; Greig, A.; Drye, T. Temporal and spatial variations in nitrogen dioxide concentrations across an urban landscape: Cambridge, UK. *Environ. Monitoring and Assessment* **1998**, *52*, 65-82.
22. Pinto, J.P.; Lefohn, A.S.; Shadwick, D.S. Spatial variability of $PM_{2.5}$ in urban areas in the United States. *J. Air & Waste Manage. Assoc.* **2004**, *54*, 440-449.
23. Butler, A.J.; Andrew, M.S.; Russell, A.G. Daily sampling of $PM_{2.5}$ in Atlanta: results of the first year of the assessement of spatial aerosol composition in Atlanta study. *J. Geophys. Res.* **2003**, *108*, 8415-8426.
24. Chow, J.C.; Watson, J.G.; Crow, D.; Lowenthal, D.H.; Merrifield, T. Comparison of IMPROVE and NIOSH carbon measurements. *Aerosol Sci. Technol.* **2001**, *34*, 23-34.
25. Cressie, N. *Statistics for Spatial Data.* Wiley and Sons: New York, 1993.
26.
27. Wilson, W.E.; Chow, J.C.; Claiborn, C.; Fusheng, W.; Engelbrecht, J.; Watson, J.G. Monitoring of particulate matter outdoors. *Chemosphere* **2002**, *49*, 1009-1043.
28. Jeong, C.; Hopke, P.K.; Kim, E.; Lee, D. The comparison between thermal-optical transmittance elemental carbon and aethelometer black carbon measured at multiple monitoring sites. *Atmos. Environ.* **2004**, *38*, 5193-5204.

Table 1. Average values and correlation coefficients of ambient air pollutants, 99-02.

pollutant measure	station	distance	avg value	R
SO_2	JS – SEARCH	0 km	18.3 / 5.9 ppb	
1-hr max / 24-hr avg	GT	1.5	13.6 / 3.9	0.68 / 0.72
	CA	8.3	11.1 / 3.1	0.67 / 0.70
	St	58.6	11.5 / 2.6	0.15 / 0.24
	Yo	60.6	9.9 / 2.7	0.10 / 0.31
CO	JS – SEARCH	0 km	1.31 / 0.53 ppm	
1-hr max / 24-hr avg	RR	11.5	1.63 / 0.80	0.61 / 0.64
	DT	16.8	1.41 / 0.63	0.74 / 0.73
	Yo	60.6	0.26 / 0.19	0.18 / 0.24
NO_x	JS – SEARCH	0 km	135 / 45.2 ppb	
1-hr max / 24-hr avg	GT	1.5	134 / 46.2	0.87 / 0.93
	SD	15.3	174 / 59.1	0.79 / 0.84
	Tu	20.4	68.1 / 25.2	0.79 / 0.84
	Co	36.8	27.5 / 10.3	0.63 / 0.62
	Yo	60.6	13.7 / 5.5	0.20 / 0.22
O_3	JS – SEARCH	0 km	64.2 / 54.4 / 30.4 ppb	
1-hr max / 8-hr max /	CA	8.3	64.2 / 55.0 / 31.3	0.97 / 0.98 / 0.94
24-hr avg, Apr-Oct	SD	15.3	60.5 / 50.9 / 25.6	0.96 / 0.97 / 0.93
	Co	36.8	61.5 / 53.4 / 29.5	0.85 / 0.88 / 0.86
	Yo	60.6	66.4 / 60.0 / 46.5	0.88 / 0.89 / 0.80
$PM_{2.5}$ mass	JS – SEARCH	0 km	17.4 $\mu g/m^3$	
24-hr avg	ERS	5.2	18.0	0.80
	SD	15.3	17.7	0.79
	DHC	18.9	18.4	0.82
	FS8 – 3rd day	3.2	20.1	0.70
	EPHC – 3rd day	17.9	18.3	0.80
	FP – 3rd day	18.7	18.2	0.87
	Ke – 3rd day	31.6	18.0	0.73
	Yo – 3rd day	60.6	15.9	0.64
	JS – TEOM	0	16.2	0.93
	FM – TEOM	7.8	19.2	0.83
	SD – TEOM	15.3	18.1	0.88
	Tu – TEOM	20.4	19.8	0.80
	Yo – SEARCH	60.6	13.4	0.84

Table 2. Average values and correlation coefficients of $PM_{2.5}$ components, March 1999 – August 2000.

$PM_{2.5}$ component	station	distance	avg value	R
sulfate	JS – SEARCH	0 km	5.5 $\mu g/m^3$	
	FM – ASACA	7.8	5.2	0.87
	SD – ASACA	15.3	5.3	0.90
	Tu – ASACA	20.4	5.2	0.92
	Yo – SEARCH	60.6	5.6	0.94
nitrate	JS – SEARCH	0 km	1.0 $\mu g/m^3$	
	FM – ASACA	7.8	0.8	0.71
	SD – ASACA	15.3	0.7	0.82
	Tu – ASACA	20.4	1.0	0.86
	Yo – SEARCH	60.6	0.8	0.61
ammonium	JS – SEARCH	0 km	2.1 $\mu g/m^3$	
	FM – ASACA	7.8	1.9	0.85
	SD – ASACA	15.3	1.8	0.90
	Tu – ASACA	20.4	2.0	0.83
	Yo – SEARCH	60.6	2.6	0.78
elemental carbon	JS – SEARCH	0 km	1.8 $\mu g/m^3$	
	FM – ASACA	7.8	1.3	0.52
	SD – ASACA	15.3	1.6	0.57
	Tu – ASACA	20.4	1.2	0.39
	Yo – SEARCH	60.6	0.7	0.43
organic carbon	JS – SEARCH	0 km	4.4 $\mu g/m^3$	
	FM – ASACA	7.8	4.0	0.49
	SD – ASACA	15.3	4.3	0.49
	Tu – ASACA	20.4	4.1	0.46
	Yo – SEARCH	60.6	3.6	0.63

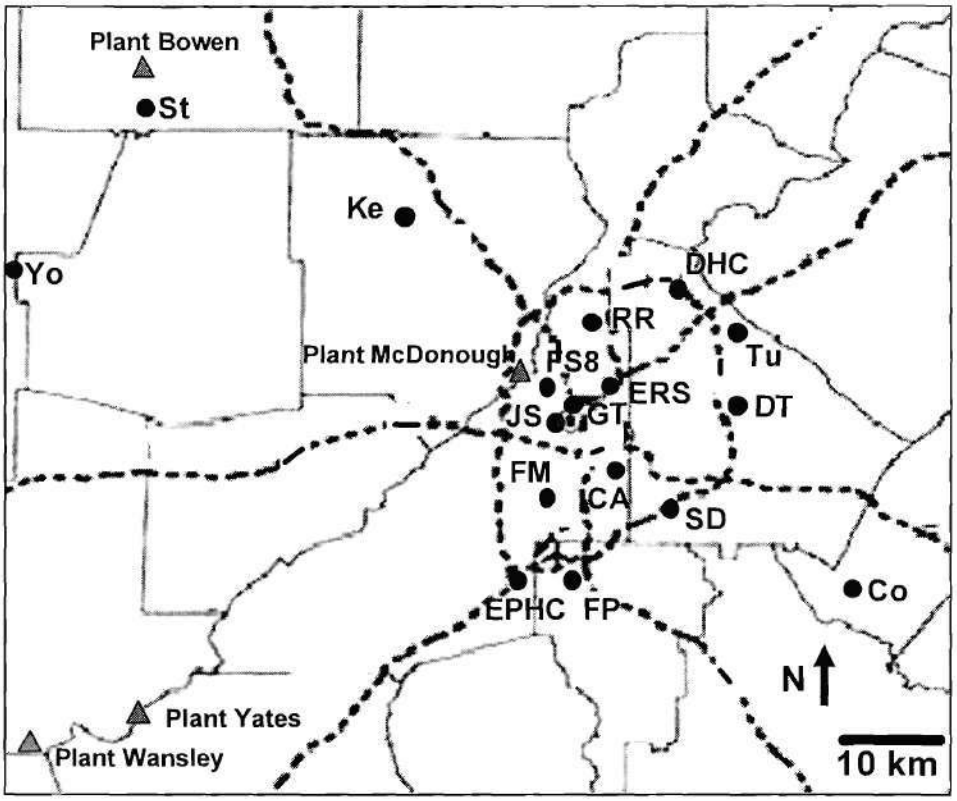

Figure 1. Locations of Atlanta metropolitan area ambient air quality monitoring sites (circles) and coal-fired power plants (triangles). Total area shown: 100 km x 100 km. County boundaries (solid lines) and interstate highways (dashed lines) are also shown.

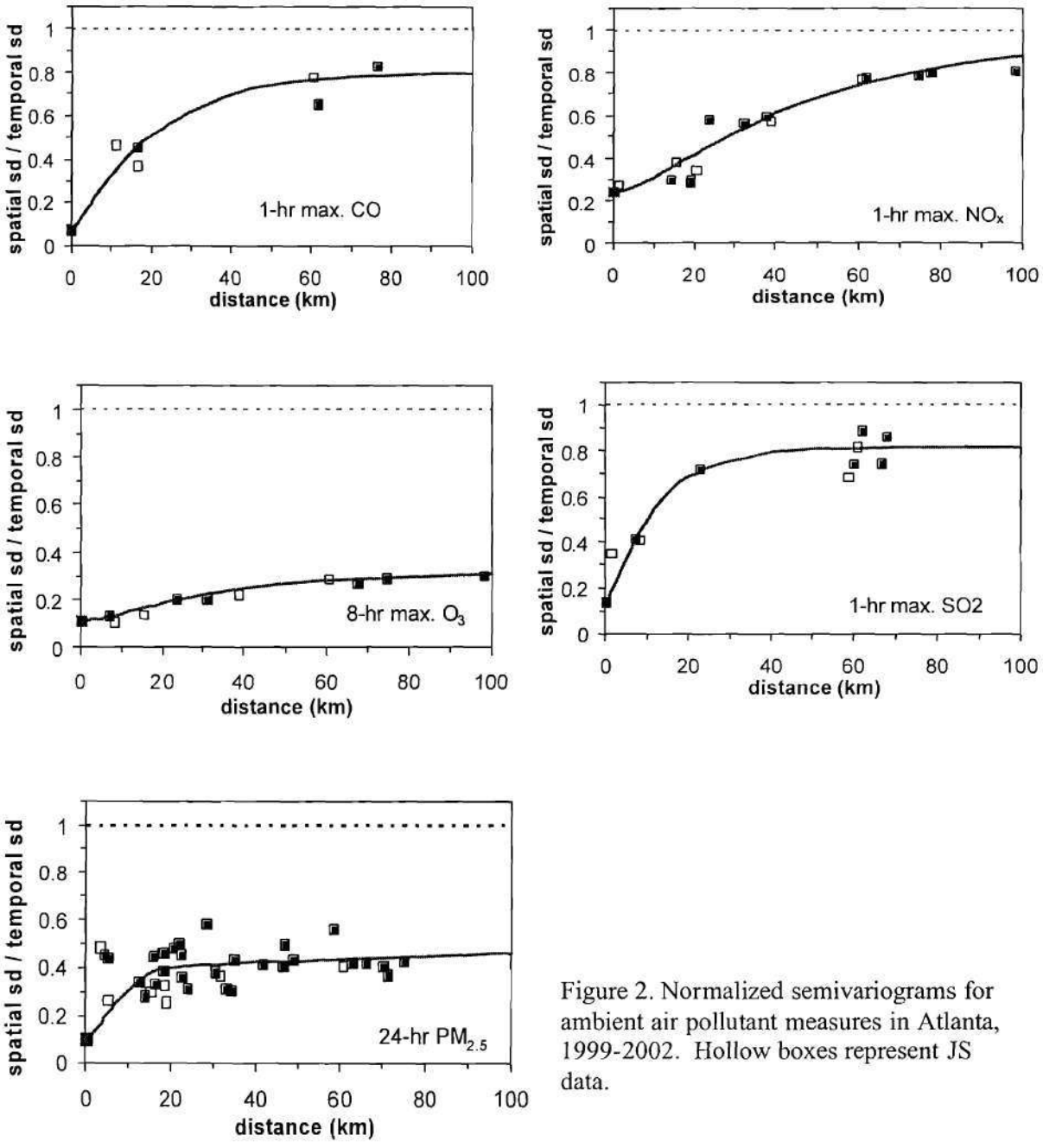

Figure 2. Normalized semivariograms for ambient air pollutant measures in Atlanta, 1999-2002. Hollow boxes represent JS data.

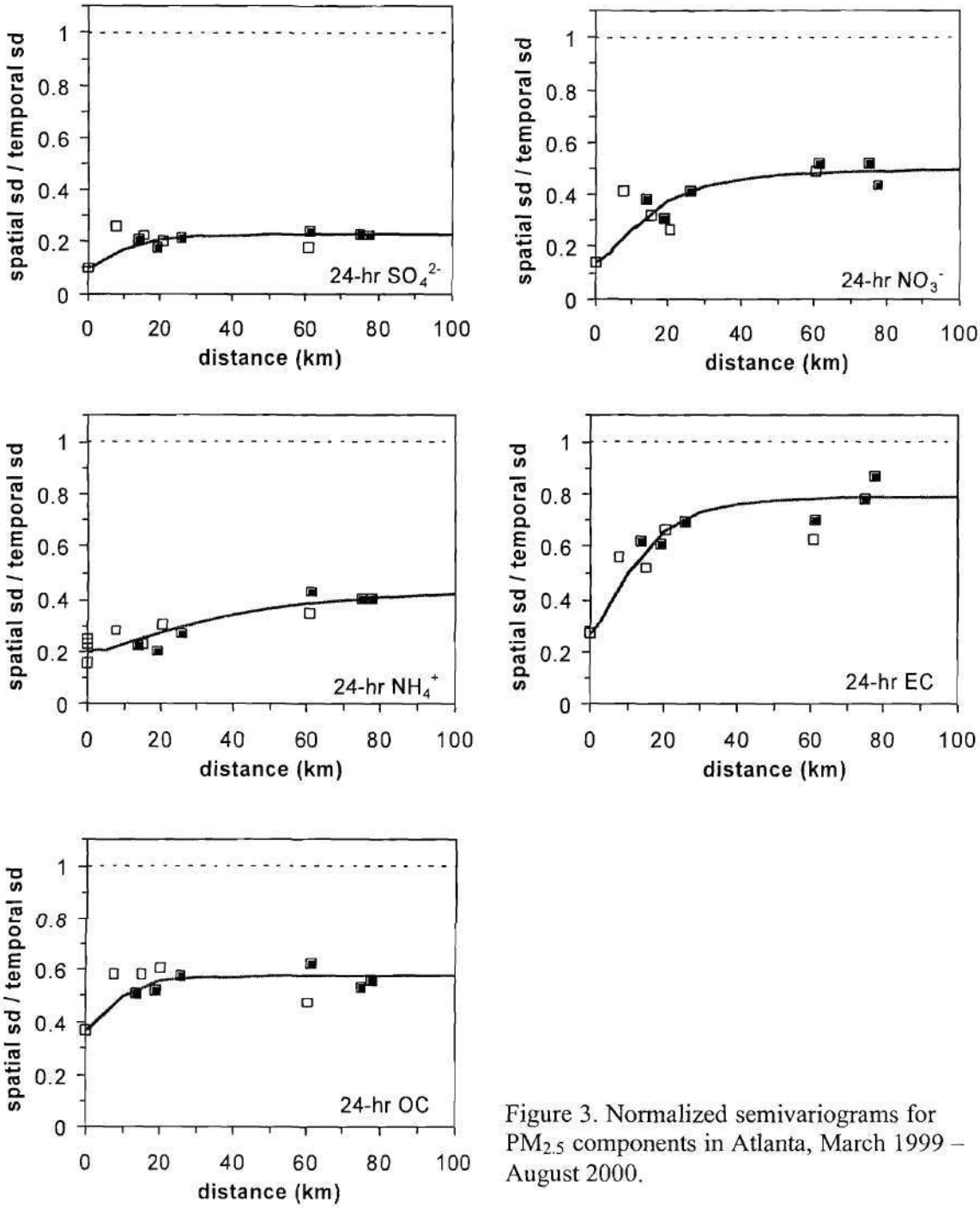

Figure 3. Normalized semivariograms for $PM_{2.5}$ components in Atlanta, March 1999 – August 2000.

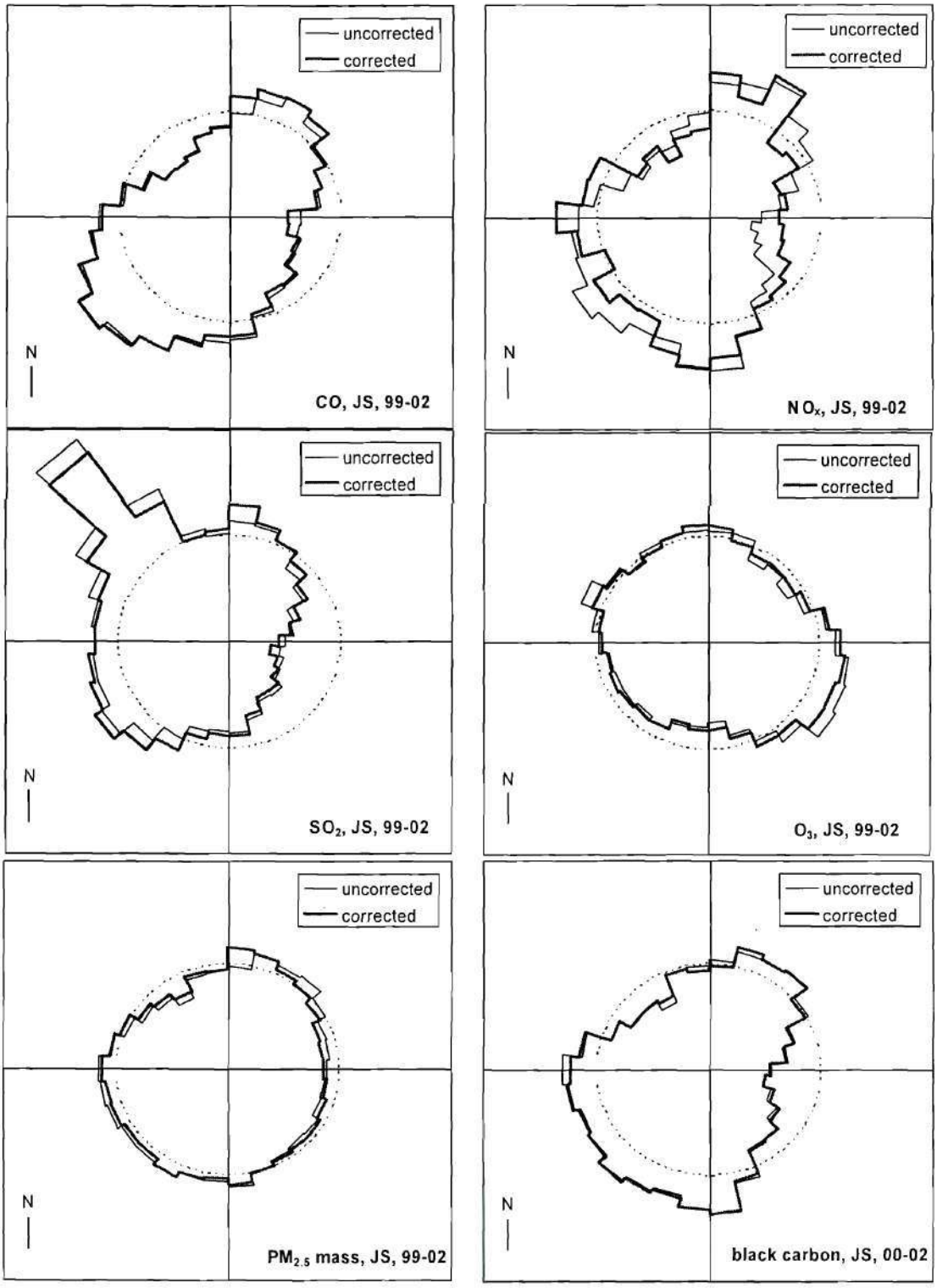

Figure 4. Wind rose plots for air pollutants measured hourly at Jefferson St. monitoring station, 1999-2002. Seasonal and diurnal effects are removed from corrected curves. Full scale is twice average value.

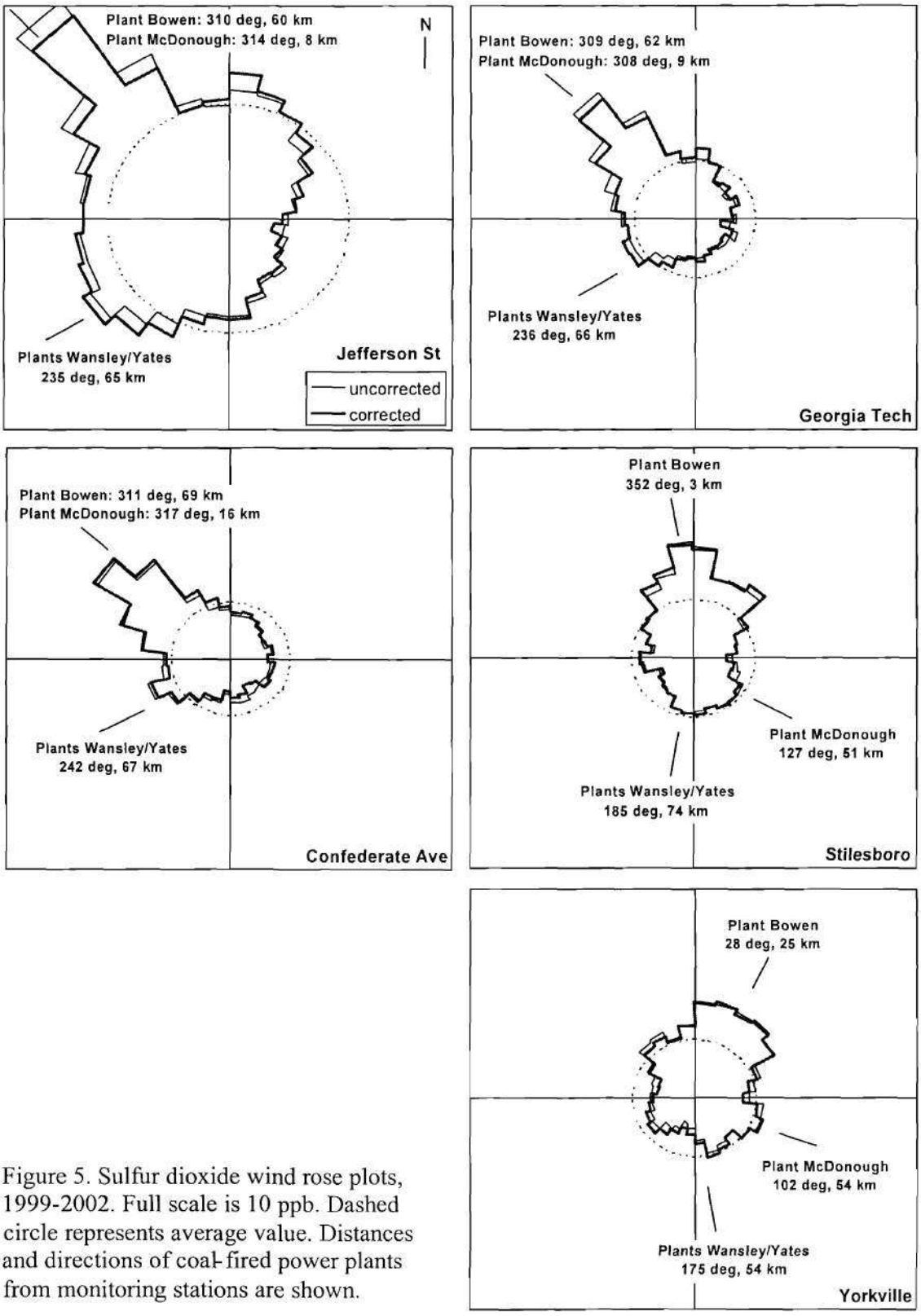

Figure 5. Sulfur dioxide wind rose plots, 1999-2002. Full scale is 10 ppb. Dashed circle represents average value. Distances and directions of coal-fired power plants from monitoring stations are shown.

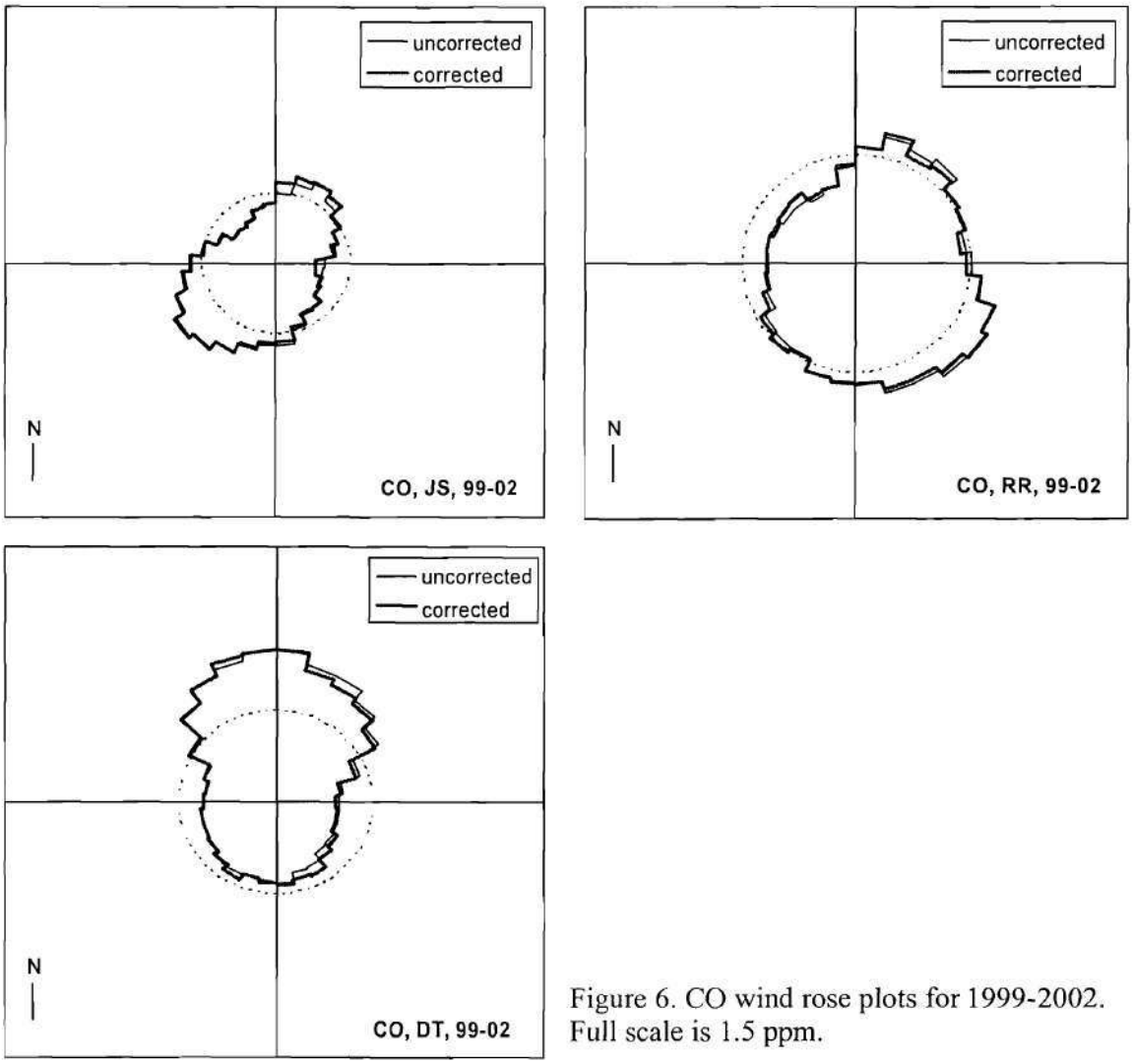

Figure 6. CO wind rose plots for 1999-2002. Full scale is 1.5 ppm.

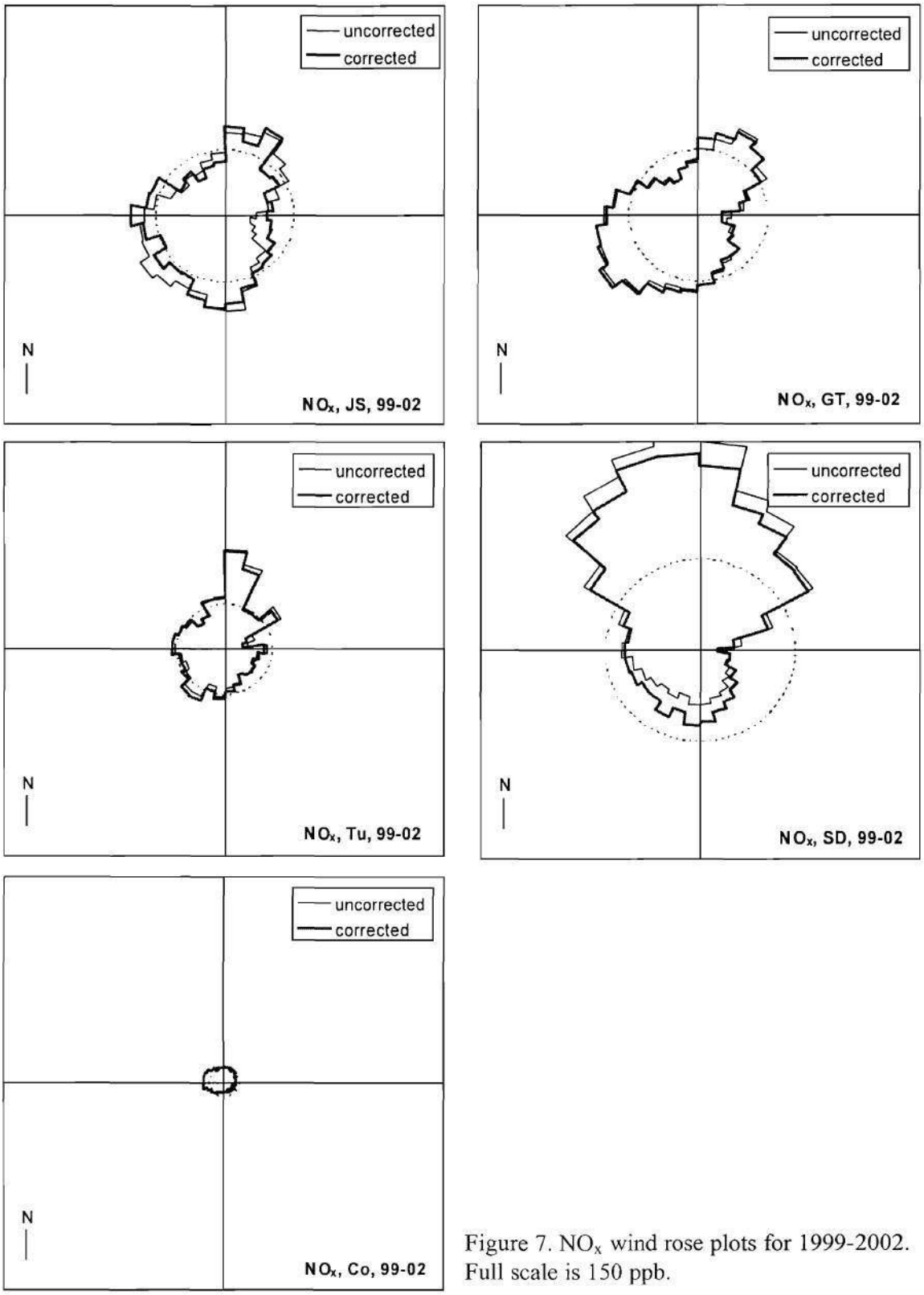

Figure 7. NO_x wind rose plots for 1999-2002. Full scale is 150 ppb.

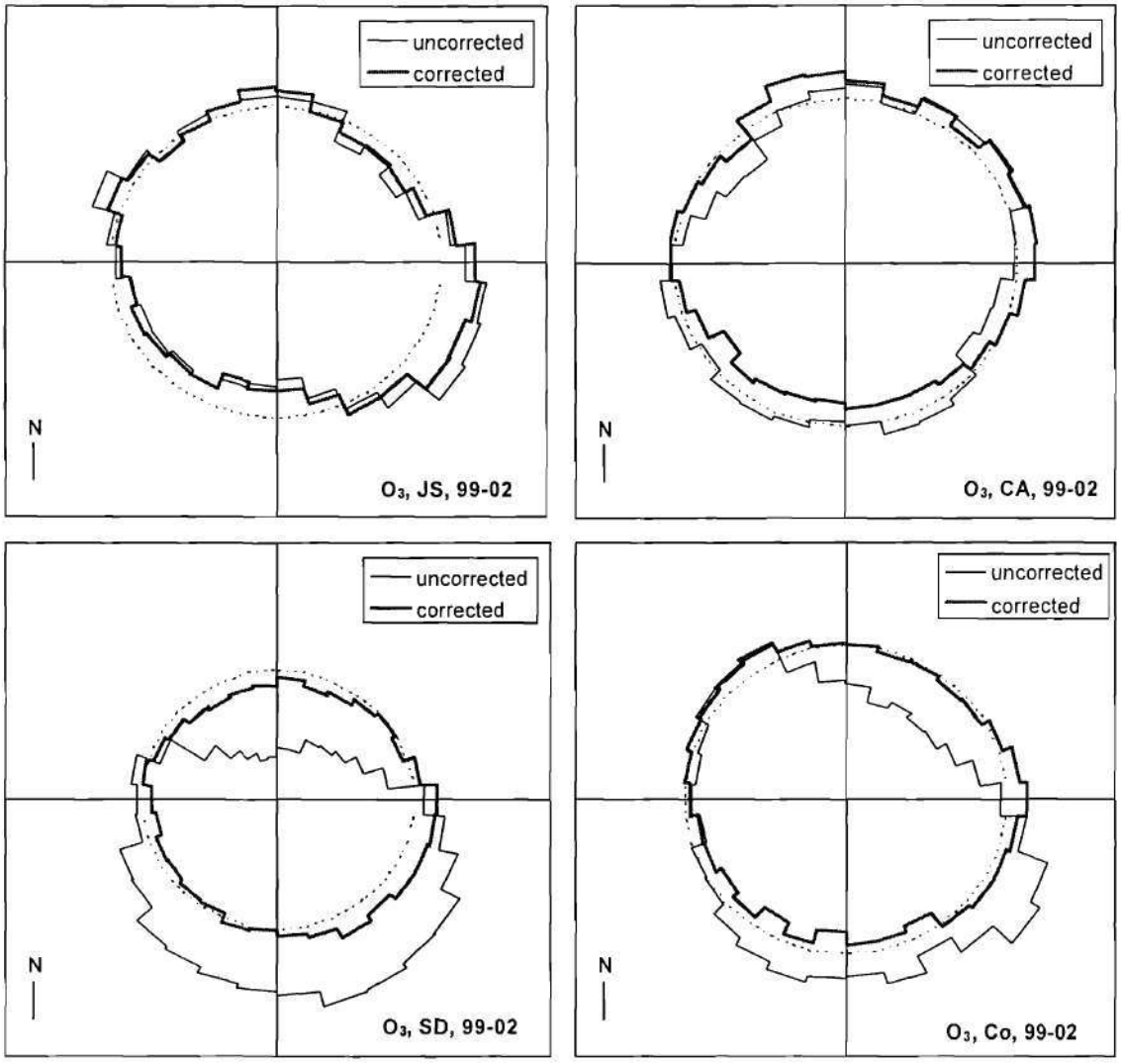

Figure 8. Ozone wind rose plots for April through October, 1999-2002. Full scale is 50 ppb.

Appendix B:

Optimization Based Source-Apportionment of PM2.5 Incorporating Gas-to-Particle Ratios

Manuscript submitted to *Environmental Science & Technology*

Optimization Based Source-Apportionment of $PM_{2.5}$ Incorporating Gas-to-Particle Ratios

Amit Marmur*, Alper Unal, Armistead G. Russell, and James A. Mulholland
School of Civil and Environmental Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0512
* corresponding author. E-mail: amit.marmur@ce.gatech.edu. Phone: 404-385-4565.

Abstract
This paper presents a new approach for $PM_{2.5}$ source-apportionment using source-indicative SO_2/EC, CO/EC and NO_x/EC ratios, in addition to the commonly used particulate-phase source-profiles. The solution is based on a global-optimization mechanism, minimizing the normalized-mean-square-error between apportioned and ambient levels of $PM_{2.5}$ components, while adhering to constraints representing the ambient SO_2/EC, CO/EC and NO_x/EC ratios. Results indicate that this technique was able to reliably identify the contribution of gasoline-vehicles to ambient $PM_{2.5}$ levels, not identified by traditional particulate-phase source-apportionment methods (due to collinearity). This contribution averaged at roughly 20% of the identifiable primary emission sources. Furthermore, the technique offered here was able to accurately identify specific power-plant fumigation events, and therefore accurately characterize the (small) direct contribution of coal-fired power-plants to ambient $PM_{2.5}$ levels, in contrast to low-resolution (over-predicted) estimates generated by traditional source-apportionment techniques.

Keywords: CMB, optimization, source-apportionment, $PM_{2.5}$, two-phase-source-profile.

Background
Chemical Mass Balance (CMB) receptor models are a common tool for apportioning of ambient levels of pollutants (mainly particulate matter) among the major contributing sources. CMB combines the chemical and physical characteristics of particles or gases measured at sources and receptors, to quantify the source contributions to the receptor. The quantification is based on the solution to linear equations that express each receptor's ambient chemical concentration as a linear sum of products of source-profile abundances and source contributions (US-EPA 2001; US-EPA, 1998), as expressed by Equation 1. The source profile abundances ($f_{i,j}$, the mass fraction of a chemical in the emissions from each source type) and the receptor concentrations (C_i), serve as input data to the CMB model. The output consists of the contribution of each source category (Sj) to the measured concentration of different species at the receptor.

$$C_i = \sum_{j=1}^{n} f_{i,j} S_j \qquad (Eq.1)$$

Where:
C_i = ambient concentration of chemical specie i ($\mu g\ m^{-3}$);
$f_{i,j}$ = fraction of specie i in emissions from source j;
Sj = *contribution (source-strength) of source* j ($\mu g\ m^{-3}$);

CMB is applicable to multi-specie data sets, the most common of which are chemically characterized particulate matter (PM), but some studies have also used CMB techniques to apportion volatile organic compounds (VOCs) (US-EPA, 1998; Lin et al., 1993). It has been widely used to quantitatively identify source contributions to ambient air pollutants at receptor sites (US-EPA, 1998). The source apportionment process is a very useful tool for air quality analysis and management, fingering out which are the source categories to be addressed, when trying to improve air-quality at specific areas of interest.

CMB models are based on the following assumptions (US-EPA, 1998):

1. Compositions of source emissions are constant over the period of ambient and source sampling.
2. Chemical species do not react with each other, i.e., they add linearly.
3. All sources with a potential for significantly contributing to the receptor have been identified and have had their emissions characterized.
4. The source compositions are linearly independent of each other.
5. The number of sources or source categories is less that or equal to the number of chemical species.
6. Measurement uncertainties are random, uncorrelated, and normally distributed.

Of these, the major assumption limiting the ability of CMB models to fully characterize the all the major sources is the linear independence of source profiles. For apportionment of $PM_{2.5}$ (particulate matter with smaller than 2.5μm), source profiles including major ions (SO_4^{-2}, NO_3^-, NH_4^+, Cl^-), elemental and organic carbon fractions (EC, OC) and trace metals are typically used. Some source categories share relatively similar (collinear) profiles (i.e. meat-cooking vs. vegetative burning, diesel vs. gasoline vehicles), limiting the ability of CMB to accurately and consistently apportion the PM mass between those sources. To address this issue, recent source-apportionment studies make use of speciated organic compounds ("organic markers"), as opposed to total OC, to further discriminate between sources (Zheng et al., 2002; Schauer et al., 2001; Schauer et al., 1999). When speciated ambient OC data, along with speciated source-profiles are available, this technique can allow higher resolution in the source apportionment process. However, speciated ambient OC data are not yet commonly available, especially not on a daily basis. As a result, organic-tracer studies are usually limited by a relatively small number of ambient measurements and by detection-level issues, which leads to lumping of daily samples into weekly or monthly averages (Zheng et al., 2002).

Enhanced $PM_{2.5}$ source-apportionment using gas-phase data

Here we offer a novel method of $PM_{2.5}$ source-apportionment, which incorporates gas-phase data, in the form of emissions of SO_2, CO and NO_x, in addition to the commonly-used $PM_{2.5}$ source-profiles. Gas-phase data can further assist in identifying sources, as sources that may have fairly similar $PM_{2.5}$ emissions, may have significantly different gas emissions. Of note are lower CO and higher NO_x emissions in diesel vehicles, compared to gas vehicles; high SO_2 emissions in coal power-plants compared to other sources; and lower NO_x emissions from vegetative burning compared to mobile-sources. Incorporating such information may allow to better differentiate between sources. Several studies have shown the increased resolution-potential in source-apportionment of two-phase receptor models (Lin et al., 1993; Wadden et al., 1991; McKee et al., 1990). Applying a two-phase receptor model for PM_{10} and non-methane-hydro-

carbons (NMHC) has shown to significantly reduce the collinearity problem (Lin et al., 1993). A study dealing with decay-adjusted receptor modeling has shown small improvements in the agreement between CMB-predicted and observed concentrations of individual VOCs, but did not significantly change the estimated emissions contributions (Lin and Milford, 1994). These studies made use of two-phase source profiles, in which the profile included the fractional composition of both PM and gas phase data (VOC, NO_x etc.), in a single profile. The profile was normalized to the sum of total PM and gases used to describe the source. When gas-phase data is incorporated in such a manner, the apportionment process becomes an apportionment of the total mass by which the profiles are normalized. For many sources of $PM_{2.5}$, the fraction of $PM_{2.5}$ emissions is much smaller that the gas phase emissions. Data from the national emission inventory for the USA (US-EPA, 1999) indicate that only about 0.6% of the total mass emissions from coal power-plants are $PM_{2.5}$, the remaining and major part being gases (SO_2, NO_x and CO). Similar trends are true for other major sources, such as mobile sources and combustion processes. Therefore, a source-apportionment process that includes two-phase source profiles is in fact mainly apportioning gas phase pollutants, being driven by the major components of the emissions. In such case, minor errors in the ratio of apportioned-to-ambient levels in the gas phase, may lead to extreme errors (over or under prediction) in the particulate phase.

To avoid inaccuracies evolving from the use of two-phase source profiles, we suggest using ratios of SO_2/EC, CO/EC and NO_x/EC as additional information assisting the apportionment process. The source profiles remain based only on the particulate matter emissions, but additional criteria based on the above ratios are used. In fact, this information adds three more equations, used as constraints, to the apportionment process, based on the same principles as in equation 1. The ambient SO_2/EC ratio can then be expressed as:

$$\frac{[SO_2]}{[EC]} = \frac{\sum_{j=1}^{n}\left(\frac{SO_2}{EC}\right)_j S_j}{\sum_{j=1}^{m} S_j} \qquad (Eq.2)$$

Where:

$\frac{[SO_2]}{[EC]}$ = ambient SO_2/EC ratio ($\mu g\ m^{-3} / \mu g\ m^{-3}$);

$\left(\frac{SO_2}{EC}\right)_j$ = SO_2/EC ratio at emissions from source j ($\mu g\ m^{-3} / \mu g\ m^{-3}$);

Sj = contribution (source-strength) of source j ($\mu g\ m^{-3}$);

n = total number of sources;

m = number of sources emitting SO_2 and/or EC ($m=n$);

Similar equations can be expressed also for the CO/EC and NO_x/EC ratios. The choice of EC as the denominator for the gas-to-particulate ratio was based on EC being the only major components of $PM_{2.5}$ which is entirely a primary pollutant. For partially secondary pollutants (such as OC), the gas-to-particulate ratio is more subject to variability, due to chemical transformations from source to receptor. A choice of a trace-level primary pollutant for the denominator would have a fairly high uncertainty associated with it. Even the choice of EC is

subject to some level of uncertainty, at the source or during transport (different deposition rates for the gases compared to EC). For that reason, the goal here is not to match the ambient and source-oriented gas/EC ratios, but rather to find an optimum solution based on the particulate-phase data, which adheres to somewhat more flexible constraints on the gaseous side. Traditional particulate-phase source-apportionment does not take that information into account.

Use of optimization techniques for source-apportionment

A large variety of quantitative decision problems in the applied sciences, engineering and economics can be described by constrained optimization models. In these models, the best decision is sought that satisfies all stated feasibility constraints and maximizes (or minimizes) the value of a given objective function. The general mathematical form of these models is summarized by (Pinter, 1996):

- max f(x)
- a = x = b
- g(x) = 0

Where :

x = a real n-vector (to describe feasible decisions)
a, b = finite, component-wise vector bounds imposed on x
f(x) = a continuous function (to describe the model objective)
g(x) = a continuous vector function (to describe the model constraints; the inequality is interpreted component-wise).

The objective of global optimization is to find the very best solution of nonlinear decision models, in the possible presence of multiple locally optimal solutions. The program system LGO, Lipschitz(-Continuous) Global Optimizer, assists in the formulation and solution of the broad class of global optimization problems described by the model form above, under minimal added analytical assumptions (Pinter, 1996). LGO integrates a suite of robust and efficient global and local scope solvers. These include: global adaptive partition and search (branch-and-bound); adaptive global random search; local (convex) unconstrained optimization; and local (convex) constrained optimization. The LGO implementation of these methods does not require derivative information. Their operations are based exclusively on the computation of the objective and constraint function values, at algorithmically selected search points. LGO has is capable of solving global optimization problems with up to hundreds of decision variables and constraints (Pinter, 1996). The list of LGO application areas includes: extremal energy (potential function) models in physical, chemical, and biological modeling; facility location and service allocation (distribution) problems; model fitting to empirical data: identification, calibration and verification procedures; risk analysis and management, and other (potentially nonconvex) stochastic decision problems; robust product or mixture design e.g., in chemical and processing industries; solution of systems of nonlinear equations and inequalities and more (Pinter, 1996).

Here, LGO was applied to describe and quantify the sources contributing to ambient levels of particulate matter. In practice, LGO was applied to solve the set of equations represented by equation 1 (22 equations, for 4 ions, 2 carbon fractions, and 16 trace metals). The solution was set subject to the constraints represented by equation 2 (three constraints, for SO_2, CO and NO_x).

The objective function chosen, which was minimized during the solution process, was the normalized-mean-square error (NMSE), defined as:

$$NMSE = \frac{\sum_{i=1}^{k}(C_i - \sum_{j=1}^{n} f_{i,j} S_j)^2}{\sum_{i=1}^{k}(C_i \sum_{j=1}^{n} f_{i,j} S_j)} \quad (Eq.3)$$

Where:
k = total number of species;

NMSE has a range of 0=NMSE=8, 0 meaning perfect agreement in value between modeled and ambient values. A NMSE value of 0.5 represents a factor of two, on the average, between the two sets of data. For receptor modeling applications, the goal is to minimize NMSE, achieving near-zero values, representing a mass balance closure (a one-to-one ratio between the sum of apportioned mass and ambient measured concentration for each of the species). The correlation coefficient cannot serve as the objective of optimization, since a correlation is not indicative of agreement in actual value (slope). Nonetheless, the optimum solution based on the minimization of the NMSE, is expected to be highly correlated with the ambient data.

Source profiles and gas-to-EC values used

The source categories used to describe the ambient levels of $PM_{2.5}$ and its components were:

- Light-Duty-Gasoline-Vehicles (LDGV)
- Heavy-Duty-Diesel-Vehicles (HDDV)
- Soil dust (SDUST)
- Vegetative-burning (BURN)
- Coal fired power plants (CFPP)
- Cement kilns (CEM)
- Meat charbroiling (COOK)
- Ammonium-sulfate (AMSULF)
- Ammonium-nitrate (AMNIT)
- Secondary OC (SECOC)

The first seven source categories are actual sources, for which measured source profiles were used. The last for categories are secondary pollutants, formed in the atmosphere. Source profiles for LDGV and HDDV were taken from real-world tunnel measurements of emissions from gasoline and diesel vehicles (HEI, 2002). Profiles for vegetative burning, power plants, cement kilns and meat cooking were taken from Big Bend Regional Aerosol Visibility and Observational (BRAVO) study (Chow et al, 2004). The soil dust profile used was from more regionally-representative measurements in Alabama (Cooper, 1981). The "profile" for the secondary pollutants was based on the molecular-weight fraction of their components. A summary of the source-profiles used in this study is given in Table 1. Secondary OC was not determined by means of source-apportionment. Instead, the EC tracer approach was used (Turpin and Huntzicker, 1995).

Table 1: Source-profiles used in the apportionment process

	LDGV[1]	HDDV[1]	SDUST[2]	BURN[3]	CFPP[3]	CEM[3]	COOK[3]	AMSULF[4]	AMNITR[4]
SO_4^{-2}	0.00000	0.00168	0.00100	0.02389	0.28743	0.31378	0.00382	0.72700	**0.00000**

NO_3^-	0.00000	0.00000	0.00100	0.00237	0.00687	0.08907	0.00269	0.00000	**0.77500**
Cl^-	0.05162	0.00489	0.00070	0.07615	0.00894	0.07121	0.00759	0.00000	**0.00000**
NH_4^+	0.00335	0.00006	0.00000	0.01648	0.01789	0.02359	0.00048	0.27300	**0.22500**
EC	0.30567	0.60845	0.00600	0.15751	0.01384	0.02956	0.10171	0.00000	**0.00000**
Pri.OC	0.26136	0.36947	0.04400	0.64406	0.27176	0.12781	0.86633	0.00000	**0.00000**
Al	0.05025	0.00000	0.09500	0.00106	0.05299	0.01064	0.00065	0.00000	**0.00000**
As	0.00000	0.00000	0.00000	0.00024	0.00003	0.00001	0.00001	0.00000	**0.00000**
Ba	0.00000	0.00000	0.00000	0.00003	0.01065	0.00040	0.00022	0.00000	**0.00000**
Br	0.00000	0.00003	0.00000	0.00083	0.00030	0.00111	0.00007	0.00000	**0.00000**
Ca	0.02583	0.00146	0.01800	0.00402	0.16555	0.17476	0.00173	0.00000	**0.00000**
Cu	0.00136	0.00018	0.00030	0.00003	0.00090	0.00024	0.00028	0.00000	**0.00000**
Fe	0.01924	0.00413	0.05300	0.00069	0.03613	0.01342	0.00315	0.00000	**0.00000**
K	0.02545	0.00108	0.00920	0.05734	0.00520	0.11592	0.00304	0.00000	**0.00000**
Mn	0.03698	0.00576	0.00160	0.00003	0.00115	0.00097	0.00026	0.00000	**0.00000**
Pb	0.00102	0.00008	0.00011	0.00003	0.00055	0.00064	0.00013	0.00000	**0.00000**
Sb	0.00000	0.00000	0.00000	0.00001	0.00011	0.00004	0.00022	0.00000	**0.00000**
Se	0.00102	0.00000	0.00000	0.00001	0.00578	0.00006	0.00000	0.00000	**0.00000**
Si	0.06823	0.00182	0.26600	0.00302	0.10694	0.04261	0.00478	0.00000	**0.00000**
Sn	0.00000	0.00000	0.00000	0.00001	0.00012	0.00005	0.00003	0.00000	**0.00000**
Ti	0.00478	0.00052	0.01000	0.00009	0.00850	0.00149	0.00007	0.00000	**0.00000**
Zn	**0.00420**	**0.00028**	**0.00009**	**0.00031**	**0.00310**	**0.00409**	**0.00036**	**0.00000**	**0.00000**

1- from HEI, 2002
2- from Cooper, 1981
3- from Chow et al., 2004
4- based on molecular-weight fractions

Gas-to-EC ratios for mobile-sources (LDGV, HDDV) were determined based on "real-world" emission estimates from tunnel studies (HEI, 2002). Ratios for vegetative burning, coal-fired power-plants, and cement plants were determined based on data from the emission inventory for the State of Georgia (reference, FAQS report?). Gas-to-$PM_{2.5}$ ratios were determined first, and then converted to gas-to-EC form, based on the fraction of EC in total $PM_{2.5}$ emissions for each source. Ratios for cooking were determined based on data from the BRAVO study (Chow et al, 2004) and other available data (McDonald et al., 2003). These ratios are reported in Table 2.

Table 2: Gas-to-$PM_{2.5}$ and Gas-to-EC ratios used as constraints (mass/mass)

Source	$SO_2/PM_{2.5}$	SO_2/EC	$CO/PM_{2.5}$	CO/EC	$NO_x/PM_{2.5}$	NO_x/EC
LDGV	6.18±0.09 [1]	20.2	-	585 (291-1200) [3]	-	**128 (79-232)** [3]
HDDV	0.820±0.006 [1]	1.35	-	20.4 (11.0-40.3) [3]	-	**74.6 (45.2-137)** [3]
Veg.Burn	0.0134±0.0004 [1]	0.0851	10.1±1.14 [1]	64.1	0.242±0.062 [1]	**1.54**
CFPP	128±29.4 [2]	9250	2.05±0.74 [2]	148	41.0±14.5 [2]	**2960**
Cement	21.5±10.1 [2]	727	0.36±0.44 [2]	12.2	18.4±23.4 [2]	**623**
Cooking	**0.504±119** [4]	**0.495**	-	**46.5±22.9** [5]	**0** [6]	**0** [6]

1- based on emission inventory data; uncertainties based on the county-level, therefore low
2- based on emission inventory data; uncertainties based on plant-level, therefore higher
3- based on tunnel studies (HEI, 2002); range based on uncertainties in both gas and EC emissions
4- based on source-profile measurements (Chow et al., 2004)
5- based on source profile measurements (McDonald et al., 2003); uncertainties based on three profiles measured
6- assuming temperature is too low for NO_x formation

The ratios reported in Table 2 demonstrate how the gas data can assist in the source apportionment process. High CO/EC ratios are indicative of gasoline-vehicles, and those can be differentiated from diesel-vehicles based the CO/EC ratio, along with the SO_2 and NO_x ratios. Vegetative-burning and cooking are characterized by low gaseous emissions, relative to the particulate phase, and can therefore be differentiated from mobile-sources. High SO_2/EC ratios are indicative of coal-fired power-plants (and to a degree also of cement-plants), and can be used to differentiate particulate power-plant emissions from soil dust (no gaseoues emission), in both of which crustal elements are abundant.

Test Case: SEARCH 25 month dataset, Jefferson St., Atlanta, Georgia

To evaluate this new approach for source-apportionment, we used the SEARCH (Southeastern Aerosol Research and Characterization) 25 month (8/98-8/00) dataset for Jefferson St. (JST) site in Atlanta, GA (Hansen et al., 2003; Kim at al., 2003). This dataset included data on total $PM_{2.5}$ mass (gravimetric measure) and its components. For the speciation of $PM_{2.5}$, a manual, filter-based, Particle Composition Monitor (PCM) was operated on a daily schedule. The PCM included three channels to collect 24 hour integrated samples for analysis of major ions, trace metals, organic and elemental carbon in $PM_{2.5}$ size range (Hansen et al., 2003; Kim at al., 2003). Ion Chromatography (IC) was used to quantify water soluble ionic species. Elemental and organic carbon collected on quartz filters were measured by Thermal Optical Reflectance (TOR). Trace metals were measured by x-ray fluorescence (XRF). Ambient values of daily SO_2, CO and NO_x were reported as well. Table 3 lists mean values and standard deviations measured at JST site, for the species used in this analysis, for the period 8/98-8/00. All values are in μgm^{-3}, except for the gas/EC ratios, which are unitless (mass/mass). For metal levels below the detection limit, half of the detection limit was reported. Organic carbon levels were split to primary and secondary organic carbon, estimated based on the EC-tracer approach (Turpin and Huntzicker, 1995). The underlying hypothesis in this approach is that because EC and primary OC often have the same sources, there is a representative ratio of OC/EC for the primary aerosol. If the measured ambient OC/EC ratio exceeds this expected value, then the additional OC can be considered to be secondary in origin. The primary OC/EC ratio was determined to be 1.8, based on the average ambient ratio of OC/EC in days of low photochemistry (using ozone as a measure for photochemistry).

Table 3: Mean, standard-deviation, minimum, maximum of ambient levels of the species used for the source-apportionment, JST site, Atlanta, GA

Specie	Mean ($\mu g/m^3$)	StDev ($\mu g/m^3$)	Min ($\mu g/m^3$)	Max ($\mu g/m^3$)
$PM_{2.5}$	19.1	8.9	1.9	**54.6**
SO_4^{-2}	5.41	3.65	0.53	**20.8**
NO_3^-	1.12	0.87	0.00	**7.49**
Cl^-	0.11	0.08	0.02	**0.83**
NH_4^+	1.99	1.32	0.15	**8.91**
EC	1.98	1.36	0.17	**11.9**
Total OC	4.46	2.21	0.66	**18.4**
Primary OC	3.07	1.71	0.30	**14.0**
Secondary OC	1.39	0.96	0.00	**6.47**
Al	1.61E-02	4.52E-02	6.16E-03	**9.00E-01**
As	1.42E-03	1.35E-03	5.05E-04	**1.51E-02**
Ba	1.81E-02	8.01E-03	1.45E-02	**5.69E-02**
Br	4.04E-03	7.97E-03	2.60E-04	**2.07E-01**
Ca	5.37E-02	4.48E-02	4.04E-03	**5.02E-01**
Cu	3.70E-03	4.57E-03	6.15E-04	**4.19E-02**
Fe	8.92E-02	7.45E-02	5.34E-03	**1.05E+00**
K	6.51E-02	5.86E-02	6.37E-03	**8.27E-01**
Mn	1.91E-03	1.54E-03	4.00E-04	**1.31E-02**
Pb	6.40E-03	7.49E-03	1.17E-03	**7.83E-02**
Sb	3.34E-03	4.40E-03	2.13E-03	**1.07E-01**
Se	1.32E-03	1.26E-03	3.50E-04	**1.01E-02**
Si	1.12E-01	1.15E-01	1.05E-02	**1.83E+00**
Sn	4.32E-03	1.92E-03	3.53E-03	**1.72E-02**
Ti	4.78E-03	4.38E-03	2.14E-03	**5.46E-02**
Zn	1.63E-02	1.61E-02	4.23E-04	**2.11E-01**
SO_2/EC (mass ratio)	10.6	11.8	0.1	**157**
CO/EC(mass ratio)	371	187	92	**2385**
NO_x/EC (mass ratio)	**55.3**	**26.1**	**4.3**	**271**

Results and discussion

Source apportionment was performed on the SEARCH 25 month dataset using three different techniques. First, CMB8 (US-EPA 2001; US-EPA 1998) was used, using the source-profiles in Table 1. Then, the LGO global optimizer (Pinter, 1996) was used without the gas-to-particle ratios. Finally, the LGO global optimizer was utilized again, this time adding constraints based on the ambient and source-specific gas-to-EC ratios. For both LGO scenarios, NMSE based on all $PM_{2.5}$ components was minimized, and an additional constraint was set, so that NMSE based on the trace metals alone (without ions or carbon data) would be limited by a value of 0.5. This, since the overall NMSE was driven mainly by the ions and carbon fractions.

Figure 1 shows the average source-contributions, based on the entire 25 month dataset (average of 762 daily values), using these three techniques. The data are also presented in Table 4. Results from all three techniques indicate that roughly a half of the ambient $PM_{2.5}$ is of secondary origin. This is indicated also by the ambient data, assuming most of the sulfate, nitrate and ammonia are of secondary origin. Daily secondary OC levels were estimated by the EC-tracer approach in all cases, and are therefore similar. Of more interest is the apportionment of the primary fraction, hence primary OC, EC and trace metals. Comparing CMB8 to the optimization approach (without gas data), both techniques indicate a significant contribution of diesel-vehicles (2.2 and 2.9 $\mu g\ m^{-3}$ respectively), cooking (2.3 and 2.0 $\mu g\ m^{-3}$ respectively) and vegetative burning processes (0.6 and 0.4 $\mu g\ m^{-3}$ respectively). CMB8 also indicates significant soil-dust (0.8 $\mu g\ m^{-3}$) and direct coal power-plant contributions (0.8 $\mu g\ m^{-3}$) (sulfate levels are a secondary

contribution of power plants to ambient $PM_{2.5}$ levels contribution). The soil-dust contribution using the optimization technique was much lower (0.2 $\mu g\ m^{-3}$) than that given by CMB8. Table 4 shows also the ratio of modeled-concentration to ambient-concentration for each of the modeled species. The data in Table 4 seem to indicate that CMB8 consistently over-predicted the levels of the major crustal-elements, hence its soil-dust contribution might be over-estimated. The optimization technique over-predicted the crustal elements as well, but to a lesser degree. The sum of contributions of coal power-plants and cement plants based on the optimization method was roughly equivalent to the power-plant contribution based on CMB8. Both techniques show no contribution of gasoline-vehicles to ambient $PM_{2.5}$ levels (CMB8 eliminated the LDGV category entirely, due to collinearity).

When adding the gas/EC constraints to the optimization process, a somewhat different solution was found. The two most dominant differences were the larger contribution of gasoline-vehicles (from a near-zero contribution) and the near-elimination of direct contributions from power-plants. The gasoline-vehicle contribution was driven by the relatively high ambient CO/EC ratios, which cannot be explained unless a fraction of the ratio was originated from gasoline-vehicles. Of note is the fact that the total contribution of mobile-sources (sum of LDGV and HDDV) did not change much (2.9 compared to 2.8 μgm^{-3}), but the diesel-contribution declined, due to the increased contribution of gasoline-vehicles. The near-elimination of the direct coal power-plant contribution was due to the low ambient SO_2/EC ratios. These ratios suggest that power-plant plumes are not a major source of primary $PM_{2.5}$. The cement plant contribution was nearly eliminated as well, also due to the low ambient SO_2/EC ratios. The soil-dust category was further decreased (0.1 $\mu g\ m^{-3}$), since the model was driven more by gas emitters. Cooking and vegetative-burning contributions did not change much.

A useful way of evaluating the reliability of source-contributions calculated by any source-apportionment model is to calculate the correlations between the daily contributions of the different sources, and the daily ambient levels of the different species. Such correlations (R values) are given in Table 5. Based on these data it can be stated that:

1. LDGV: A major difference is observed in the correlations when using the gas-data, compared to more traditional source-apportionment. Contributions from gasoline vehicles were not well correlated for the optimization without gas data case, as the source contribution was near-zero (no correlations reported for CMB8, given a zero-contribution). On the other hand, the optimization based source-apportionment with gas-data produced contributions that were well correlated with the major components of the LDGV source-profile. These include EC, OC, Ca, Fe, Mn, and Si.
2. HDDV: The three source-apportionment techniques used here generated high correlations with the major components of diesel emissions (EC, OC, Fe, Mn).
3. BURN: A unique marker for vegetative burning is potassium. CMB8 and the optimization with gas-data techniques generated high correlations with potassium, indicative of vegetative-burning.
4. SDUST: soil-dust is characterized by a high abundance of crustal-elements, such as Al, Ca, Fe Si and Ti. Results from all three techniques used here were correlated with these crustal elements. CMB8 was driven mainly by Fe, while the two optimization methods were driven mainly by Si, which is the major component of soil-dust.
5. CFPP: A major difference is observed in the correlations when using the gas-data, compared to more traditional source-apportionment. CMB8 and the optimization without gas-data were driven mainly by EC, OC, Fe and Ca levels, all abundant in direct

emissions from coal-fired power-plants, but also in other sources such as gasoline-vehicles and cement-kilns, and soil-dust to a degree. But when the SO_2/EC constraint was used (optimization with gas-data), selenium, a unique marker for coal-fired power-plants, was the element driving the apportionment process. The correlation was relatively low (0.31), likely due to the small and infrequent source-impact. Using the SO_2/EC ratio it was possible to detect actual days in fumigation of the power-plant plume occurred, and to apportion part of the ambient $PM_{2.5}$ to the power-plant category. That apportioned fraction (relatively low) was mostly correlated with Se. When the SO_2/EC ratio was not used, it seemed difficult to differentiate between direct power-plant emissions and other sources, hence the relatively high power-plant contribution.

6. CEM: Cement kilns emissions were correlated mainly with crustal elements. CMB8 was driven mainly by Al, Ti and K, while the optimization without gas-data by Ca, K, Si and Ti. Using the gas-data, cement emissions were correlated with Ca only.
7. COOK: Data from Table 4 indicate meat-charbroiling as major source for $PM_{2.5}$, with contributions averaging at about 2 μgm^{-3}. It is difficult to validate these numbers without using specific organic-markers for cooking, such as cholesterol. Using a non-speciated OC profile, the major components of meat-charbroiling emissions are OC (87%) and EC (10%), with no unique markers. Meat-charbroiling processes also cannot be easily characterized by the gas/EC ratios. All three techniques indicated high correlations with EC and OC. One possible explanation for the high meat-charbroiling contribution could be that it actually represents secondary OC, given that the profile is comprised almost solely of OC. But in that case, we would not expect a high correlation with EC and primary OC, as indicated in Table 5. In addition, the contribution of secondary OC was estimated separately using the EC-tracer approach, and was deducted form the ambient OC levels used for the apportionment. That contribution averaged at 1.4 $\mu g\ m^{-3}$, roughly 30% of the total OC. If the cooking contribution were actually additional secondary OC, that would have indicated an average secondary OC fraction of more than 80% of the total OC, which seems too high.
8. SECOC: Daily secondary OC levels were estimated using the EC-tracer approach (Turpin and Huntzicker, 1995). As indicated in Table 5, secondary OC levels were not well correlated with any of the noted species. Such low of correlations are expected, as the assumption is that the sources for secondary OC are different from those of EC and primary OC. For comparison, the correlation between ambient levels of total OC and EC is 0.8, as indicated by the JST dataset. Nevertheless, the lack of correlation between secondary OC and EC, by itself, cannot validate the magnitude of daily contributions estimated here.

Figure 1: Source-contributions to $PM_{2.5}$ levels measured at JST site, Atlanta, GA, using CMB8, optimization technique without gas-data, and optimization technique with gas-data

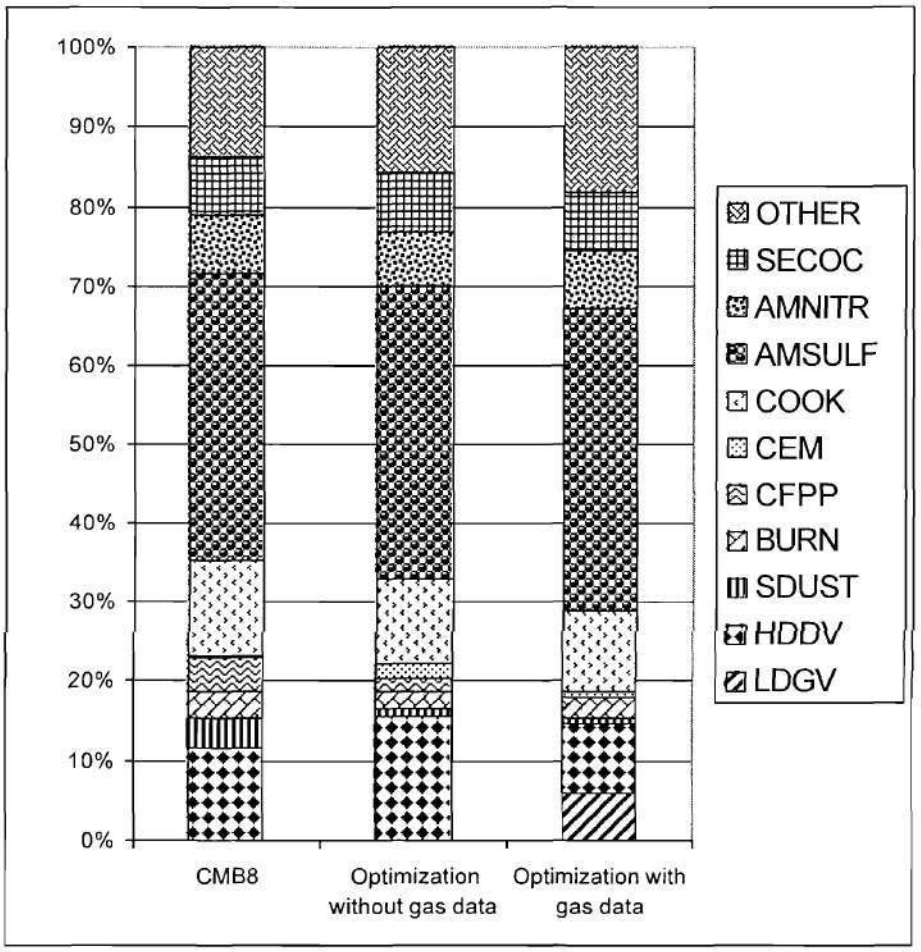
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
CMB8
Optimization without gas data
Optimization with gas data
OTHER
SECOC
AMNITR
AMSULF
COOK
CEM
CFPP
BURN
SDUST
HDDV
LDGV

Table 4: Source-contributions to $PM_{2.5}$ levels measured at JST site, Atlanta, GA, using CMB8, optimization technique without gas-data, and optimization technique with gas-data. Also reported are the correlation (R), NMSE, % total mass, and modeled-to-measured ratios

	CMB8	**Optimization without gas data**	Optimization with gas data
	Mean (StDev)	Mean (StDev)	**Mean (StDev)**
R [1]	0.990 (0.179)	0.995 (0.014)	**0.982 (0.035)**
NMSE $PM_{2.5}$ [1]	-	0.024 (0.140)	**0.046 (0.106)**
NMSE metals [2]	-	0.447 (0.153)	**0.487 (0.173)**
NMSE gas [3]	-	79.2 (120.6)	**1.60 (1.72)**
% total mass [4]	80.7 (16.3)	78.8 (18.8)	**76.1 (17.5)**
LDGV ($\mu g/m^3$)	0.00 (0.00)	0.02 (0.15)	**1.14 (0.70)**
HDDV ($\mu g/m^3$)	2.20 (1.70)	2.94 (2.30)	**1.67 (1.84)**
SDUST ($\mu g/m^3$)	0.75 (0.85)	0.20 (0.39)	**0.11 (0.36)**
BURN ($\mu g/m^3$)	0.61 (0.78)	0.43 (0.59)	**0.52 (0.77)**
CFPP ($\mu g/m^3$)	0.81 (0.63)	0.30 (0.35)	**0.004 (0.01)**
CEM ($\mu g/m^3$)	0.04 (0.14)	0.34 (0.37)	**0.11 (0.13)**
COOK ($\mu g/m^3$)	2.30 (1.61)	2.05 (1.77)	**1.95 (1.53)**
AMSULF ($\mu g/m^3$)	6.95 (4.91)	7.08 (4.93)	**7.31 (4.93)**
AMNITR ($\mu g/m^3$)	1.42 (1.13)	1.33 (1.17)	**1.40 (1.29)**
SECOC ($\mu g/m^3$) [5]	1.39 (0.96)	1.39 (0.96)	**1.39 (0.96)**
SO4 ratio [6]	0.979 (0.022)	0.992 (0.070)	**0.999 (0.096)**
NO3 ratio [6]	1.02 (0.848)	0.994 (1.92)	**0.959 (0.535)**
Cl ratio [6]	1.07 (0.931)	1.10 (0.772)	**1.68 (1.36)**
NH4 ratio [6]	1.16 (0.154)	1.17 (0.196)	**1.22 (0.207)**
EC ratio [6]	0.857 (0.064)	1.06 (0.323)	**0.861 (0.443)**
Prim. OC ratio [6]	1.12 (0.104)	1.07 (0.386)	**0.973 (0.335)**
Al ratio [6]	11.7 (9.24)	3.99 (2.71)	**7.33 (4.00)**
As ratio [6]	0.21 (0.27)	0.15 (0.21)	**0.16 (0.26)**
Ba ratio [6]	0.55 (0.38)	0.23 (0.22)	**0.03 (0.02)**
Br ratio [6]	0.35 (0.35)	0.42 (0.51)	**0.27 (0.31)**
Ca ratio [6]	3.52 (1.88)	2.64 (1.16)	**1.29 (0.54)**
Cu ratio [6]	1.04 (0.85)	0.82 (0.68)	**1.38 (1.20)**
Fe ratio [6]	0.97 (0.01)	0.58 (0.18)	**0.55 (0.20)**
K ratio [6]	0.92 (0.48)	1.35 (0.59)	**1.36 (0.50)**
Mn ratio [6]	11.1 (9.24)	14.0 (12.3)	**41.4 (34.0)**
Pb ratio [6]	0.29 (0.25)	0.29 (0.26)	**0.51 (0.45)**
Sb ratio [6]	0.22 (0.15)	0.19 (0.15)	**0.16 (0.13)**
Se ratio [6]	5.27 (5.29)	2.38 (3.42)	**1.59 (1.40)**
Si ratio [6]	2.89 (1.54)	1.11 (0.26)	**1.26 (0.45)**
Sn ratio [6]	0.04 (0.02)	0.03 (0.02)	**0.02 (0.01)**
Ti ratio [6]	3.88 (2.54)	1.70 (0.94)	**1.99 (1.02)**
Zn ratio [6]	0.33 (0.20)	0.37 (0.32)	**0.58 (0.49)**
SO_2/EC ratio [6]	-	101 (428)	**3.20 (1.30)**
CO/EC ratio [6]	-	0.121 (0.060)	**0.471 (0.116)**
NO_x/EC ratio [6]	-	**4.80 (4.25)**	**1.36 (0.63)**

1- based on all $PM_{2.5}$ components
2- based on trace metals only
3- based on gas/EC ratios
4- % of apportioned mass to total $PM_{2.5}$
5- based on the EC-tracer approach
6- ratio of apportioned mass to ambient level (ideally would approach 1 for all species)

Table 5: Correlations (R) between source-contributions and ambient-levels (Opt.-optimization without gas data, O.G.- optimization with gas-data)

	LDGV			HDDV			BURN		
	CMB8	Opt.	O.G.	CMB8	Opt.	O.G.	CMB8	Opt.	**O.G.**
SO_4^{-2}	-	0.34	0.22	0.22	0.19	0.25	0.13	0.18	**0.29**
NO_3^-	-	0.34	0.10	0.22	0.25	0.16	0.09	0.26	**0.10**
Cl^-	-	0.05	0.01	0.12	0.15	0.08	0.22	**0.54**	**0.24**
NH_4^+	-	**0.41**	0.23	0.27	0.23	0.29	0.10	0.26	**0.26**
EC	-	**0.39**	**0.56**	**0.98**	**0.94**	**0.72**	0.21	0.28	**0.28**
Pri. OC	-	0.28	**0.56**	**0.73**	**0.73**	**0.57**	0.22	0.32	**0.33**
Al	-	0.04	0.32	0.17	0.09	0.06	0.01	0.20	**0.11**
As	-	0.34	0.22	0.33	0.33	0.28	0.18	0.09	**0.17**
Ba	-	0.04	0.17	0.11	0.08	0.05	0.11	0.10	**0.19**
Br	-	0.29	0.08	0.15	0.17	0.13	0.38	0.26	**0.29**
Ca	-	0.31	**0.56**	0.42	0.33	0.42	0.10	0.25	**0.15**
Cu	-	0.12	0.38	0.44	0.43	0.38	0.11	0.21	**0.20**
Fe	-	0.30	**0.62**	**0.72**	**0.66**	**0.57**	0.03	0.34	**0.23**
K	-	0.35	0.37	0.36	0.36	0.24	**0.82**	**0.49**	0.69
Mn	-	**0.41**	**0.50**	**0.69**	**0.63**	**0.53**	0.04	0.29	**0.19**
Pb	-	0.05	0.22	0.37	0.43	0.36	0.18	0.30	**0.23**
Sb	-	-0.05	0.00	0.06	0.07	0.09	0.24	0.03	**0.06**
Se	-	**0.52**	0.26	0.42	0.35	0.27	0.11	0.11	**0.13**
Si	-	0.24	**0.55**	0.37	0.27	0.23	0.05	0.28	**0.24**
Sn	-	-0.10	0.21	0.16	0.15	0.08	0.13	0.09	**0.06**
Ti	-	0.33	0.49	0.44	0.35	0.33	0.14	0.36	**0.29**
Zn	-	0.31	0.36	0.48	0.48	0.44	0.16	0.28	**0.19**
	SDUST			**CFPP**			CEM		
	CMB8	Opt.	O.G.	CMB8	Opt.	O.G.	CMB8	Opt.	**O.G.**
SO_4^{-2}	0.20	0.17	0.08	0.22	0.40	0.14	0.08	0.38	**0.13**
NO_3^-	0.05	-0.02	0.00	0.24	0.01	0.30	0.04	0.05	**0.15**
Cl^-	0.02	0.00	0.03	0.18	0.03	0.27	0.34	0.22	**0.06**
NH_4^+	0.23	0.18	0.07	0.22	0.36	0.19	0.08	0.35	**0.13**
EC	0.49	0.22	0.11	**0.69**	**0.39**	0.15	0.30	0.27	**0.19**
Pri. OC	0.35	0.12	0.02	**0.62**	**0.41**	0.17	0.25	0.30	**0.21**
Al	**0.77**	**0.91**	**0.94**	0.26	0.22	0.08	**0.58**	0.28	**0.10**
As	0.17	0.08	0.07	0.37	0.11	0.08	0.01	0.12	**0.16**
Ba	0.18	0.20	0.16	0.13	0.17	-0.01	0.12	0.20	**0.04**
Br	0.07	0.01	-0.01	0.22	0.10	0.09	0.20	0.04	**0.24**
Ca	0.65	**0.71**	**0.61**	0.60	**0.59**	0.15	0.21	**0.47**	0.46
Cu	0.40	0.21	0.12	0.56	0.22	0.01	0.30	0.21	**0.11**
Fe	**0.96**	**0.76**	**0.71**	**0.80**	**0.44**	0.17	0.30	0.37	**0.24**
K	0.34	0.29	0.27	0.53	**0.39**	0.16	**0.73**	**0.66**	**0.24**
Mn	**0.72**	0.52	0.48	**0.71**	0.33	0.19	0.29	0.32	**0.21**
Pb	0.26	0.09	0.02	0.53	0.07	0.20	**0.48**	0.12	**0.08**
Sb	0.02	-0.02	-0.03	0.12	0.05	0.02	-0.02	-0.01	**0.20**
Se	0.20	0.14	0.12	0.38	0.24	**0.31**	0.07	0.25	**0.32**
Si	**0.88**	**0.96**	**0.92**	0.47	**0.41**	0.13	0.29	**0.44**	**0.21**
Sn	0.13	0.10	0.04	0.19	0.11	-0.07	0.16	0.09	**0.05**
Ti	**0.79**	**0.80**	**0.81**	0.49	**0.42**	0.10	**0.52**	**0.47**	**0.15**
Zn	**0.26**	**0.14**	**0.03**	**0.68**	**0.24**	**0.14**	**0.40**	**0.16**	**0.20**

Table 5 (cont.): Correlations (R) between source-contributions and ambient-levels (Opt.- optimization without gas data, O.G.- optimization with gas-data)

	COOK			AMSULF			AMNITR			SECOC
	CMB8	Opt.	O.G.	CMB8	Opt.	O.G.	CMB8	Opt.	O.G.	**EC-tracer**
SO_4^{-2}	0.22	0.11	0.18	**1.00**	**1.00**	**0.99**	0.05	0.02	0.09	**0.27**
NO_3^-	0.23	0.21	0.17	0.04	0.06	0.06	**1.00**	**0.98**	**0.92**	**0.08**
Cl^-	0.14	0.14	0.11	0.06	0.07	0.07	0.38	0.37	0.33	**0.07**
NH_4^+	0.24	0.12	0.18	**0.94**	**0.95**	**0.94**	0.23	0.23	0.28	**0.26**
EC	**0.76**	**0.53**	**0.53**	0.19	0.21	0.23	0.23	0.22	0.21	**0.39**
Pri. OC	**0.95**	**0.74**	**0.76**	0.20	0.21	0.23	0.26	0.24	0.23	**0.31**
Al	0.02	-0.04	-0.05	0.00	0.00	0.01	-0.04	-0.05	-0.05	**0.00**
As	0.28	0.27	0.20	-0.02	0.00	0.01	0.18	0.17	0.16	**0.23**
Ba	0.07	0.03	0.10	0.17	0.17	0.18	-0.01	-0.01	0.00	**0.08**
Br	0.43	0.17	0.30	0.00	0.03	0.03	0.11	0.10	0.28	**0.09**
Ca	0.36	0.17	0.31	0.26	0.26	0.28	0.06	0.04	0.05	**0.15**
Cu	0.43	0.36	0.31	0.07	0.09	0.10	0.11	0.12	0.11	**0.20**
Fe	**0.63**	0.45	0.33	0.21	0.24	0.25	0.13	0.14	0.14	**0.24**
K	0.43	0.23	0.25	0.12	0.13	0.15	0.11	0.09	0.09	**0.33**
Mn	0.51	0.35	0.33	0.18	0.20	0.21	0.19	0.20	0.22	**0.23**
Pb	0.42	0.44	0.29	-0.04	0.00	-0.01	0.23	0.26	0.24	**0.25**
Sb	0.17	0.11	0.24	-0.06	-0.04	-0.03	0.01	0.00	0.02	**0.01**
Se	0.21	0.09	0.13	0.37	0.38	0.39	0.16	0.13	0.12	**0.22**
Si	0.28	0.10	0.10	0.24	0.24	0.25	-0.06	-0.07	-0.07	**0.15**
Sn	0.13	0.08	0.12	0.00	0.01	0.02	0.09	0.08	0.08	**0.11**
Ti	0.34	0.16	0.18	0.26	0.26	0.28	-0.03	-0.04	-0.04	**0.13**
Zn	**0.46**	**0.41**	**0.32**	**0.00**	**0.03**	**0.04**	**0.23**	**0.23**	**0.21**	**0.35**

Summary

A new approach for source-apportionment based on global-optimization of the normalized-mean-square-error, subject to gas/EC ratios, was presented. The approach makes use of the additional source-identification potential of SO_2/EC, CO/EC and NO_x/EC ratios, to drive the $PM_{2.5}$ source-apportionment process through a solution that would adhere to constraints representing the ambient SO_2/EC, CO/EC and NO_x/EC ratios. Results indicated that this technique was able to reliably identify the contribution of gasoline-vehicles to ambient $PM_{2.5}$ levels, not identified by traditional particulate-phase source apportionment methods. This contribution averaged at 1.1 $\mu g\ m^{-3}$, roughly 20% of the identifiable primary emission sources. Furthermore, this technique was able to accurately identify specific power-plant fumigation events, and therefore accurately characterize the direct contribution of coal-fired power-plants to ambient $PM_{2.5}$ levels. This contribution was negligible, in contrast to findings from traditional source-apportionment. Contributions attributed to direct power-plant emissions using the traditional method (without the SO_2/EC) ratio may just as well be indicative to other sources, such as cement kilns and gasoline-vehicles. It is unclear whether the relatively large (2 $\mu g\ m^{-3}$ roughly) contribution attributed the meat-charbroiling, by all methods, is truly representative.

References

1. Chow, J.C., Watson, J.G., Kuhns, H., Etyemezian, V., Lowenthal, D.H., Crow, D., Kohl, S.D., Engelbrecht, J.P., and Green, M.C. (2004), Source Profiles for Industrial, Mobile, and Area Sources in the Big Bend Regional Aerosol Visibility and Observational study, Chemosphere, 54:185-208.
2. Cooper, J.A. (1981), Determination of Source Contributions to Fine and Coarse Suspended Particulate Levels in Petersville, Alabama. Report to Tennessee Valley Authority by NEA, Inc.
3. FAQS EI report
4. Hansen, D.A.; Edgerton, E.S.; Hartsell, B.E.; Jansen, J.J.; Kandasamy, N.; Hidy, G.M.; Blanchard, C.L. *Journal of the Air and Waste Management Association*, 2003, 53, 1460-1471.
5. Health Effects Institute (HEI) (2002), Emissions from Diesel and Gasoline Engines Measured in Highway Tunnels, Research Report # 107.
6. Kim, E., Hopke, P.K., and Edgerton, E.S. (2003), Source Identification of Atlanta Aerosol by Positive Matrix Factorization, Journal of the Air & Waste Management Association, 53:731-739.
7. Lin, C. and Milford, J.B. (1994), Decay-Adjusted Chemical Mass Balance Receptor Modeling for Volatile Organic Compounds, Atmospheric Environment, 28: 3261-3276.
8. Lin, J., Scheff, P.A. and Wadden, R.A (1993), Development of a Two-Phase Receptor Model for NMHC and PM_{10} Air Pollution Sources in Chicago, presented at the 86th annual meeting & Exhibition of the Air & Waste Management Association, Denver, Colorado.
9. McDonald, J.D., Zielinska, B., Fujita, E.M., Sagebiel, J.C., Chow, J.C., and Watson, J.G. (2003), Emissions from Charbroiling and Grilling of Chicken and Beef, Journal of the Air & Waste Management Association, 53:185-194.
10. McKee, G.A., Wadden, R.A. and Scheff, P.A. (1990), Development of a Two-Phase Chemical Mass Balance Receptor Model, presented at the 83rd annual meeting & Exhibition of the Air & Waste Management Association, Pittsburgh, Pennsylvania.
11. Pinter, J.D. (1996), Global Optimization in Action, Kluwer Academic Publishers, The Netherlands.
12. Schauer, J.J., Kleeman, M.J., Cass, G.R., Simoneit, A.T. (1999), Measurement of Emissions from Air Pollution Sources. 2. C_1 through C_{30} Organic Compounds from Medium Duty Diesel Trucks, Environmental Science and Technology, 33:1578-1587.
13. Schauer, J.J., Kleeman, M.J., Cass, G.R., Simoneit, A.T. (2001), Measurement of Emissions from Air Pollution Sources. 3. C_1-C_{29} Organic Compounds from Fireplace Combustion of Wood, Environmental Science and Technology, 35:1716-1728.
14. Seinfeld, J.H. and Pandis, S.N. (1998), Atmospheric Chemistry and Physics, John Wiley & Sons, Inc.,
15. Turpin, B.J., and Huntzicker, J.J. (1995), Identification of Secondary Organic Aerosol Episodes and Quantification of Primary and Secondary Organic Aerosol Concentrations During SCAQS, Atmospheric Environment, 29:3527-3544.
16. US-EPA (1998), CMB8 Application and Validation Protocol for $PM_{2.5}$ and VOC, EPA-454/R-98-XXX, Office of Air Quality, Planning and Standards, Research Triangle Park, NC 27711.

17. US-EPA (1999), National Emission Inventory (NEI): Air Pollutant Emission Trend (www.epa.gov/ttn/chief/trends/).
18. US-EPA (2001), CMB8 User's Manual, EPA-454/R-01-XXX, Office of Air Quality, Planning and Standards, Research Triangle Park, NC 27711.
19. Wadden, R.A., Scheff, P.A., Lin. J., Lee, H., Keil, C., Graf-Teterycz, J., Keehan, K., Kenski, D., Milz, S., Holsen, T.M. and Khalili, N. (1991), Two Phase Receptor Modeling, presented at the 84th annual meeting & Exhibition of the Air & Waste Management Association, Vancouver, British Columbia.
20. Zheng, M., Cass, G.R., Schauer, J.J. and Edgerton, E.S., (2002), Source Apportionment of $PM_{2.5}$ in the Southeastern United States Using Solvent-Extractable Organic Compounds as Tracers, Environmental Science and Technology, 36:2361-2371.